

# A multivariate invariance principle for modified wild bootstrap methods with an application to unit root testing

Citation for published version (APA):

Smeeke, S., & Urbain, J. R. Y. J. (2014). *A multivariate invariance principle for modified wild bootstrap methods with an application to unit root testing*. Maastricht University, Graduate School of Business and Economics. GSBE Research Memoranda No. 008 <https://doi.org/10.26481/umagsb.2014008>

## Document status and date:

Published: 01/01/2014

## DOI:

[10.26481/umagsb.2014008](https://doi.org/10.26481/umagsb.2014008)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

Stephan Smeekes,  
Jean-Pierre Urbain

**A Multivariate Invariance  
Principle for Modified Wild  
Bootstrap Methods with an  
Application to Unit Root  
Testing**

RM/14/008

**GSBE**

Maastricht University School of Business and Economics  
Graduate School of Business and Economics

P.O. Box 616  
NL- 6200 MD Maastricht  
The Netherlands

# A MULTIVARIATE INVARIANCE PRINCIPLE FOR MODIFIED WILD BOOTSTRAP METHODS WITH AN APPLICATION TO UNIT ROOT TESTING \*

Stephan Smeekes

Jean-Pierre Urbain

Department of Quantitative Economics

Maastricht University

February 25, 2014

## Abstract

In this paper we consider several modified wild bootstrap methods that, additionally to heteroskedasticity, can take dependence into account. The modified wild bootstrap methods are shown to correctly replicate an invariance principle for multivariate time series that are characterized by general forms of unconditional heteroskedasticity, or nonstationary volatility, as well as dependence within and between different elements of the time series. The invariance principle is then applied to derive the asymptotic validity of the wild bootstrap methods for unit root testing in a multivariate setting. The resulting tests, which can also be interpreted as panel unit root tests, are valid under more general assumptions than most current tests used in the literature. A simulation study is performed to evaluate the small sample properties of the bootstrap unit root tests.

*JEL Classification:* C15, C32.

*Keywords:* invariance principle, modified wild bootstrap, multivariate time series, unconditional heteroskedasticity, unit root test.

---

\*Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. E-mail: [S.Smeekes@maastrichtuniversity.nl](mailto:S.Smeekes@maastrichtuniversity.nl), [J.Urbain@maastrichtuniversity.nl](mailto:J.Urbain@maastrichtuniversity.nl). The first author would like to thank the Netherlands Organisation for Scientific Research (NWO) for financial support. We also thank Eric Beutner for helpful discussions. Previous versions of this paper entitled “Unit Root Testing using Modified Wild Bootstrap Methods” were presented at the OxMetrics User Conference 2011 in Maastricht, the Statistische Woche 2011 in Leipzig, the 6th CSDA Conference on Computational and Financial Econometrics in Oviedo, the 2012 (EC)<sup>2</sup> Conference in Maastricht, the 2013 Joint Statistical Meetings in Montreal, and the Workshop on Recent Developments in Bootstrap Methods for Time Series Data in Copenhagen, as well as seminars at Maastricht University, the Tinbergen Institute and Deakin University. We are grateful to conference and seminar participants for helpful comments and suggestions. All remaining errors are our own.

# 1 Introduction

Many macroeconomic and financial time series exhibit changes in unconditional volatility; in particular, there is a large body of empirical evidence for a general decline in unconditional volatility in the shocks driving macroeconomic time series since the eighties in what is often called the Great Moderation; see, among others, Kim and Nelson (1999), McConnell and Perez Quiros (2000), Blanchard and Simon (2001), Stock and Watson (2003), Sensier and Van Dijk (2004) and references therein. After the occurrence of the 2008 credit crisis and the following financial turmoil and disruption of economic activity, many have argued that the general decline in unconditional volatility has been reversed and a sharp rise has taken place; in other words, that the great moderation has ended, see for example Taylor (2011). Whether the recent increase is really permanent or not, the recent crisis only reinforces the observation that unconditional volatilities have not remained constant over time.

It is therefore crucial to account for such unconditional heteroskedasticity when performing econometric inference. Nowadays many methods are available for this purpose. One particularly successful method for robustifying inference with respect to unconditional heteroskedasticity is the wild bootstrap (cf. Davidson and Flachaire, 2008). However, the wild bootstrap assumes independence of the innovations and is therefore not robust to serial correlation in the data. In the presence of autocorrelation one can add a sieve, or recoloring, component to the wild bootstrap, as advocated by Cavaliere and Taylor (2008, 2009a,b) for the purpose of unit root testing. While this is generally a satisfactory solution to handle autocorrelation in univariate time series, or small multivariate systems, this is not the case for large multivariate systems or panel data. A multivariate autoregressive sieve bootstrap cannot be applied if the dimension of the system is too large, and Smeekes and Urbain (2014) have shown that a univariate autoregressive sieve bootstrap cannot correctly replicate dynamic dependence structures among the different units in the system. However, many recent developments in macroeconometric modeling involve large multivariate systems, or panel data, of macroeconomic time series typically collected over a number of countries; see for example Canova and Ciccarelli (2013) and the references therein. As such time series can be subject to the kind of volatility changes described above, appropriate methods are needed for dealing with unconditional heteroskedasticity in large multivariate systems.

Recently several methods have been developed for dealing with heteroskedasticity in multivariate time series and panel data. Particularly worth mentioning in the context of panel unit root testing is the approach of Demetrescu and Hanck (2012), that uses IV estimators, and the approach by Westerlund (2014), who uses the cross-sectional dimension to consistently estimate the time-varying volatility and adjusts the series accordingly. These methods however do not utilize the bootstrap, even though in univariate time series the wild bootstrap provides a nearly unrivaled success in delivering robust inference with respect to heteroskedasticity; not only in the unit root setting but also for stationary time series (see e.g. Gonçalves

and Kilian, 2004, 2007).

In this paper we therefore try to preserve the good properties of the wild bootstrap while adapting it for applications in multivariate time series. This we achieve by investigating modifications of the wild bootstrap that allow one to simultaneously deal with serial dependence as well as heteroskedasticity in multivariate time series. These modified wild bootstrap methods can be seen as “block” versions of the original wild bootstrap, and therefore provide a heteroskedasticity-robust analogue of the block bootstrap. The nonparametric block structure makes the methods particularly suitable for application in large multivariate systems such as the ones encountered in typical macroeconomic applications where the dimension is too large to estimate (semi-)parametric models for the serial dependence. While standard block bootstrap methods have a similar advantage in large multivariate systems with respect to the capturing of the serial dependence, those methods are not appropriate in the presence of unconditional heteroskedasticity. As an additional advantage over traditional block bootstrap methods, the modified wild bootstrap methods considered here are considerably easier to apply if not all time series in the system are of equal length.

Some of the modifications to the wild bootstrap considered here have been proposed by Shao (2010, 2011) in a different context, also see Leucht and Neumann (2013) and Doukhan, Lang, Leucht, and Neumann (2014) for other applications of similar modifications. In this paper we establish a multivariate invariance principle for the modified wild bootstrap methods applied to time series with general dependence structures and volatility processes. In particular we demonstrate that the asymptotic behavior of the partial sums of the modified wild bootstrap time series correctly mimics the asymptotic behaviour of the original time series. This powerful result implies that the modified wild bootstrap methods yield asymptotically valid inference for multivariate time series in the presence of weak dependence and unconditional heteroskedasticity.

We next apply the established bootstrap invariance principle to multivariate, or panel, unit root testing. We propose unit root tests based on the modified wild bootstrap methods and demonstrate these tests are asymptotically correctly sized and consistent. Their finite sample performance is also investigated by the use of Monte Carlo simulations and contrasted to tests based on a traditional block bootstrap as well as an autoregressive sieve bootstrap, both of which are invalid (though for different reasons).

The structure of the paper is as follows. In Section 2 we derive the multivariate invariance principle for the modified wild bootstrap methods. The invariance principle is applied to unit root testing in Section 3. The finite-sample properties of the modified wild bootstrap unit root tests are explored through Monte Carlo simulation in Section 4. Section 5 concludes. All proofs are contained in the Appendix.

Finally, a word on notation. Weak convergence (convergence in distribution) is denoted by  $\overset{d}{\rightarrow}$ , convergence in probability is denoted by  $\overset{p}{\rightarrow}$ . Bootstrap quantities, conditional on the

original sample, are denoted with a superscript ‘\*’. Bootstrap weak convergence in probability (cf. Giné and Zinn, 1990) is denoted by ‘ $\xrightarrow{d^*}_p$ ’.  $\|A\|$  applied to a matrix or vector  $A$  denotes its norm  $\|A\| = \sqrt{\text{tr}(A'A)}$ .  $\lfloor x \rfloor$  is the largest integer smaller than or equal to  $x$ .  $W(r)$  denotes a standard  $N$ -dimensional Brownian motion.  $D_{\mathbb{R}^{N \times N}}[0, 1]$  denotes the space of  $N \times N$  matrices of càdlàg functions on  $[0, 1]$ .

## 2 Invariance Principle

### 2.1 The Model and Preliminary Results

We consider a multivariate process  $\{u_t\}_{t=1}^T$ , where

$$u_t = (u_{1,t}, \dots, u_{N,t})' = \sigma_t v_t, \quad (1)$$

where  $\{\sigma_t\}$  governs the shape of the volatility of  $\{u_t\}$ , while  $\{v_t\}$  is a weakly dependent stationary stochastic process. Specifically, we assume  $\sigma_t$  and  $v_t$  satisfy the following conditions.

#### Assumption 1.

- (i)  $v_t$  is an  $N$ -dimensional strictly stationary strong mixing process, with  $\mathbb{E} v_t = 0$ ,  $\mathbb{E} v_t v_t' = I_N$ ,  $\mathbb{E} \|v_t\|^{(2+\delta)} < \infty$  for some  $\delta > 0$ , and with strong mixing coefficients  $\alpha(m)$  satisfying  $\sum_{m=0}^{\infty} \alpha(m)^{1/p-1/(2+\delta)} < \infty$  for some  $2 < p < 2 + \delta$ . Furthermore,  $\Lambda = \lim_{T \rightarrow \infty} \mathbb{E} \left( \sum_{t=1}^T v_t \right) \left( \sum_{t=1}^T v_t \right)'$  is a positive semi-definite matrix.
- (ii)  $\sigma_t$  is a deterministic  $N \times N$  matrix such that  $\sigma_t = \sigma(t/T)$ , where  $\sigma(\cdot) \in D_{\mathbb{R}^{N \times N}}[0, 1]$  is piecewise Lipschitz with a finite number of discontinuities and  $\Sigma(u) = \sigma(u)\sigma(u)'$  is positive definite for all  $u \in [0, 1]$ .

Assumption 1(i) is a standard mixing assumption that is commonly assumed in applications of the block bootstrap (cf. Lahiri, 2003). The assumption on the mixing coefficients is very weak and essentially only guarantees absolute summability of the autocovariances of  $\{v_t\}$  and existence of a functional central limit theorem (cf. Hansen, 1992). This very general assumption leaves the nature of the serial dependence unspecified, both within one variable and between variables. As such, dynamic interdependencies, where the lag of one variable affects the present of another variable (cf. Canova and Ciccarelli, 2013), but also common factors, often of interest in macroeconomic applications, are allowed.

Assumption 1(ii) is similar to Assumption 2 of Cavaliere, Rahbek, and Taylor (2010) and allows for a wide variety of volatility processes, including single and multiple shifts, trends and smooth transitions in variance (cf. Remark 2.1 of Cavaliere et al., 2010). The requirement that  $\sigma(\cdot)$  is piecewise Lipschitz is a slight strengthening of Assumption 2 of Cavaliere et al. (2010), required by the way we impose the nonstationary volatility on the mixing process  $\{v_t\}$ ,

rather than inside a linear process as Cavaliere et al. (2010) do.<sup>1</sup> For practical purposes this stronger assumption only rules out pathological cases, and all empirically relevant processes such as the ones described above, remain admissible.

**Remark 1.** The assumption that  $\Lambda$  can have a reduced rank is not of particular interest in the  $I(0)$  world, but is important in Section 3 when we consider  $I(1)$  processes, as this assumption then allows for cointegration between variables. See Remark 9 for details.

**Remark 2.** Assumption 1(ii) allows for change in both variances and covariances, and as such, the DGP also allows for time-varying correlation. This concept is often deemed important in financial variables, for example in the context of contagion. Also, common volatility shocks are a possibility that the DGP allows for (cf. Cavaliere et al., 2010, Remark 2.3).

**Remark 3.** Assumption 1(ii) excludes stochastic volatility processes. This allows for considerable simplification of the theoretical arguments (also see Remark 2.2 in Cavaliere et al., 2010). Cavaliere and Taylor (2009b) prove the validity of univariate unit root tests based on autoregressive sieve bootstrap with a wild bootstrap component under a more general assumption that allows for stochastic volatility processes. We believe that this result can be extended to the methods developed in the current paper. However, given the high expositional cost, the formal validity of our methods under such an assumption is outside the scope of this paper. We will return to this issue in our simulation study in Section 4.

**Remark 4.** Conditional heteroskedasticity such as produced by GARCH processes is allowed under Assumption 1(i). The assumptions on stationarity, mixing and, in particular, the existence of moments do place restrictions on the type of processes allowed; for example, see Ling and McAleer (2002), Francq and Zakoïan (2010, Chapters 2 and 3) and the references cited therein for the exact conditions on the parameters of GARCH processes that are needed in order to satisfy Assumption 1(i).

As a first step of our theoretical analysis, we now present the invariance principle for  $\{u_t\}$  in Lemma 1.

**Lemma 1.** *Let  $u_t$  be generated by (1) and let Assumption 1 hold. Then, as  $T \rightarrow \infty$ ,*

$$M_T(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} u_t \xrightarrow{d} M(r), \quad (2)$$

where  $M(r) = \int_0^r \sigma(s) dB(s)$ ,  $B(r) = \lambda W(r)$  and  $\lambda \lambda' = \Lambda$ .

It is important to note that the vector partial sum process of  $\{u_t\}$  does not converge to a vector Brownian motion, but instead to the martingale  $M(\cdot)$ , which depends on both

---

<sup>1</sup>Compare to Assumption  $\mathcal{V}$  in Cavaliere (2005), who uses a similar framework as ours.

the volatility process  $\sigma(\cdot)$  and the long-run covariance matrix  $\Lambda$ . In particular the complex dependence on the volatility process  $\sigma(\cdot)$  makes achieving pivotal inference problematic. A second problem in achieving pivotal inference is correcting for the off-diagonal elements of  $\Lambda$ , in particular when  $N$  is reasonably large. As such, while one may attempt to correct the statistics of interest for these two sources of nuisance parameters, this is difficult in practice, in particular if the form of the volatility process and the nature of the dependence are unknown. We therefore do not go down this route, but instead attempt to mimic the influence of both types of nuisance parameters by capturing them with our bootstrap methods.

## 2.2 Modified Wild Bootstrap (MWB) Methods

The wild bootstrap in its original form is designed to mimic the heteroskedasticity present in the data. In particular, it generates the bootstrap sample  $\{u_t^*\}_{t=1}^T$  as

$$u_t^* = \xi_t^* u_t, \tag{3}$$

where  $\{\xi_t^*\}_{t=1}^T$  are identically distributed scalar random variables with  $\mathbb{E}^* \xi_t^* = 0$  and  $\mathbb{E}^* (\xi_t^*)^2 = 1$ . Crucially, in the standard wild bootstrap  $\{\xi_t^*\}_{t=1}^T$  are generated independently.<sup>2</sup> These conditions ensure that the unconditional heteroskedasticity is asymptotically correctly mimicked in the bootstrap sample, see e.g. Cavaliere et al. (2010, Lemma A.5) for a multivariate bootstrap invariance principle. However, while contemporaneous correlation is automatically captured by the fact that  $\{\xi_t^*\}$  is a scalar, any serial correlation present in  $u_t$  will be lost in the bootstrap. As such, the long-run covariance matrix  $\Lambda$  is not correctly replicated by the standard wild bootstrap. Therefore, unless the statistic for which the wild bootstrap is used in a specific application can be corrected for serial correlation, for example by augmenting with lags or estimating long-run covariances nonparametrically, the standard wild bootstrap yields asymptotically invalid inference. Typically, for large multivariate systems such a correction is infeasible as the number of parameters to be estimated increases too quickly. Even if asymptotically consistent corrections are available, their finite sample performance might not be sufficient to eliminate the impact of serial correlation.

Therefore it is generally advisable to modify the wild bootstrap directly in such a way that the serial correlation is mimicked by the wild bootstrap, rather than relying only on successful corrections of the statistics. While this is already important in univariate time series, in particular in small samples, it is crucial in large multivariate systems, not just in small samples, but often even asymptotically, as the construction of asymptotically pivotal statistics is generally quite difficult for potentially complex forms of dependence.

---

<sup>2</sup>This still leaves several possibilities to generate  $\{\xi_t^*\}$ . Apart from being asymptotically equivalent, in the context of unit root testing these different options only have marginal finite sample effects. Therefore the typical approach, as advocated by Cavaliere and Taylor (2008) or Shao (2010) for example, is generate them as standard normal.



One method of modification is to combine the wild bootstrap with an autoregressive sieve component. Specifically, one applies the wild bootstrap to residuals from an autoregressive model and builds the bootstrap series recursively using the estimated autoregressive parameters. We label this approach the Sieve Wild Bootstrap (SWB). For the purposes of univariate unit root testing this approach was proposed by Cavaliere and Taylor (2008, 2009a,b) and Smeeke and Taylor (2012). A multivariate version of the SWB in the context of cointegration testing can be found in Cavaliere et al. (2010), where the wild bootstrap is combined with the estimation of a vector autoregressive (VAR) model.

In multivariate time series whose dimension is too large to estimate a VAR with sufficient accuracy, the approach of Cavaliere et al. (2010) is infeasible. Instead, in such a setting one can estimate an autoregressive model for each unit individually and then apply the wild bootstrap to the residuals jointly. Smeeke and Urbain (2014) analyze such an approach for homoskedastic time series, and demonstrate that it neglects dynamic cross-correlation that is present between variables. As a result the off-diagonal elements of  $\Lambda$  are not replicated correctly by the SWB, which may lead to invalid inference.

Given the inherent problems of estimating VAR models in large dimensions and the invalidity of the SWB variant described above, we consider a modification of the wild bootstrap along a different road. Rather than applying a sieve principle we modify the wild bootstrap directly by allowing the random variables  $\{\xi_t^*\}_{t=1}^T$  to be dependent, which can be seen as a block version of the wild bootstrap. We consider three modifications here, collectively labeled as Modified Wild Bootstrap (MWB).

**Block Wild Bootstrap [BWB]** The block wild bootstrap (BWB) was proposed as block-wise wild bootstrap by Shao (2011). Define  $k = \lceil T/\ell \rceil$  and generate  $\nu_1^*, \dots, \nu_k^*$  as i.i.d.  $N(0, 1)$ . Then let

$$\xi_t^* = \xi_{(m-1)\ell+s}^* = \nu_m^*, \quad t = 1, \dots, T,$$

where  $m = \lceil t/\ell \rceil$  and  $s = t - (m-1)\ell$ . The tuning parameter  $\ell = \ell_T$  determines the block length and as such how much of the dependence should be retained. As  $\xi_t^* = \nu_1^*$  for all  $t = (m-1)\ell + 1, \dots, m\ell$ , we have that  $u_t^* = \nu_1^* \hat{u}_t$  for all  $t = 1, \dots, \ell$ ,  $u_t^* = \nu_2^* \hat{u}_t$  for all  $t = \ell + 1, \dots, 2\ell$ , and so on. Hence dependence within one block is preserved completely, but between blocks it is lost. This treatment of the dependence mirrors that of the non-overlapping block bootstrap, with block length  $\ell$ , and the BWB can therefore be seen as the wild bootstrap extension thereof.

**Dependent Wild Bootstrap [DWB]** The dependent wild bootstrap (DWB) was proposed by Shao (2010). It aims to generate  $\xi_1^*, \dots, \xi_T^*$  such that  $\text{Cov}(\xi_s^*, \xi_t^*) = K\left(\frac{s-t}{\ell}\right)$ , where  $K(\cdot)$  is a kernel function with  $K(0) = 1$ ,  $K(x) = 0$  for  $x \geq 1$ , and  $\int_{-\infty}^{\infty} K(u) e^{-iux} dx > 0$  for

all  $x \in \mathbb{R}$ . The parameter  $\ell$  again determines the extent of the dependence that is retained. As variables further than  $\ell$  units apart are independent, it can again be interpreted as a block length parameter. To implement the DWB we generate  $\xi_1^*, \dots, \xi_T^* \sim N(0, \Sigma)$ , where  $\Sigma = \left( K\left(\frac{i-j}{\ell}\right) \right)_{i,j=1}^N$ . For the choice of kernel we follow the recommendation of Shao (2010) in Remark 2.1.

Just as the BWB has a block bootstrap interpretation, the DWB can also be interpreted as a counterpart to a block bootstrap. However, while the BWB is the counterpart of the non-overlapping block bootstrap, the DWB can be interpreted as the counterpart of the moving block bootstrap (MBB) or the related tapered block bootstrap with overlapping blocks (see Shao, 2010, Section 2). As such one could say that it provides a less “coarse” version of the BWB. A related difference between the two methods is that the bootstrap variables  $\{\xi_t^*\}$  are stationary for the DWB, yet clearly not for the BWB.<sup>3</sup>

**Autoregressive Wild Bootstrap [AWB]** The third modification we consider is the autoregressive wild bootstrap (AWB). For this method generate  $\nu_1^*, \dots, \nu_T^*$  as i.i.d.  $N(0, 1 - \gamma^2)$  and let

$$\xi_t^* = \gamma \xi_{t-1}^* + \nu_t, \quad t = 2, \dots, T, \quad \xi_1^* \sim N(0, 1).$$

where  $\gamma = \gamma_T \in [0, 1)$  is a tuning parameter that controls the persistence of the process. A practitioner can select the value of  $\gamma$  itself as a tuning parameter, but as seen in Section 2.3,  $\gamma$  has to increase to unity with the sample size to achieve asymptotic validity of the bootstrap. The most straightforward way to achieve this is to write  $\gamma = \theta^{1/\ell}$ , where  $\theta$  is a fixed parameter, and  $\ell$  is the same tuning parameter as in the previous methods. This parametrization makes this method comparable to the previous two. Note that  $\text{Cov}(\xi_s^*, \xi_t^*) = \gamma^{|s-t|} = \theta^{|s-t|/\ell}$ . By defining the family of kernels  $K_\theta(x) = \theta^{|x|}$ , the AWB method can be put in the framework of the DWB, where the choice of  $\theta$  can be seen as the choice of the kernel. An important difference though is that in the DWB method the bootstrap innovations more than  $\ell$  units apart are independent. This is not true for the AWB, where bootstrap innovations only become independent asymptotically. Leucht and Neumann (2013) and Doukhan et al. (2014) consider a similar type of modification for the analysis of (strictly stationary) time series in the context of  $U$ - and  $V$ -statistics and the empirical process, respectively.

**Remark 5.** Conceptually all three modified wild bootstrap methods described above (and by construction the standard, independent, wild bootstrap as well) can be seen as special cases of one general wild bootstrap method that is defined by the equations  $\mathbb{E}^* \xi_t = 0$  and  $\mathbb{E}^* \xi_s^* \xi_t^* = \tilde{K}(s, t, \ell)$  for all  $s, t = 1, \dots, T$ . We keep treating the three methods separately however. Our main reason is that in developing the theory, it is more convenient to treat

---

<sup>3</sup>Though this clearly does not mean that for the DWB  $\{u_t^*\}$  is stationary conditional on the data.

the methods separately. Though the proofs for the three methods are similar, the BWB in particular requires a different approach than the other two methods as the variables  $\{\xi_t^*\}$  are not stationary for the BWB, in contrast to the DWB and AWB.

An additional motivation to treat the methods separately is that they all have separate interpretations. First, the BWB has as much a different interpretation as the other two methods as the non-overlapping block bootstrap has compared with the moving block bootstrap. Second, we maintain the DWB in its current form as Shao (2010) proposed it in the same way, though for a different purpose. Third, while similar to the DWB, the AWB in its current form has a very straightforward and in a time series context more natural interpretation than the DWB. We therefore keep treating it separately as well.

**Remark 6.** The choice of the tuning parameter  $\ell$  is a trade off between capturing more dependence (by increasing  $\ell$ ) and allowing for more variation in the bootstrap samples which leads to a better approximation to the sampling distribution (by decreasing  $\ell$ ). This trade off is, unsurprisingly, very similar to the trade off faced when selecting the block length in the block bootstrap. Choosing an optimal block length is a topic that has had much attention in the bootstrap literature. Unfortunately, this is still an open topic, particularly in the context of unit root testing. Palm, Smeeke, and Urbain (2011) provide an overview of different methods and investigate their performance by simulation for panel unit root testing, but theoretical results are still unavailable.

An additional complication arising in the presence of time-varying volatility is that the tuning parameter  $\ell$  has a direct effect on how much of the variation in the volatility is captured. As elaborated on in Section 2.3, the MWB methods have a close correspondence to long-run variance estimation (also see Shao, 2010, Section 2), where the tuning parameter determines how many auto-covariance are used in the estimation of the long-run variances. Using a large tuning parameter however also has the effect that volatility changes are smoothed out in the bootstrap; see equation (5). Therefore, a small tuning parameter would be preferred if volatility changes are highly abrupt, highlighting the trade off with capturing serial dependence further.

Given the close resemblance to long-run variance estimation one might consider using rules for optimal bandwidth selection in the current context. However, it is unlikely that those methods are optimal in this setting, as they do not take into account the changes in volatility. To find an optimal tuning parameter selection method higher order asymptotic theory is needed, but this is complicated by the “double nonstationarity” present and therefore outside the scope of this paper.

**Remark 7.** The choice of  $\gamma$  (or equivalently  $\theta$  and  $\ell$ ) in the AWB should be seen as the choice of a tuning parameter that is not a direct reflection of the form of the serial correlation structure in the DGP (i.e. the true dependence does not have to be of AR(1) form). While it

might appear to be an intuitively logical choice to choose the tuning parameter  $\gamma$  in such a way that it matches the dependence found in the sample, at least in a univariate context, the theoretical implications of such a scheme are not directly obvious; for example, it is difficult to see how this could be matched with the required condition that  $\gamma$  has to increase to 1 with the sample size.

### 2.3 Bootstrap Invariance Principle

We now derive the multivariate invariance principle for partial sum processes obtained from modified wild bootstrap samples that allows one to establish asymptotic validity of the modified wild bootstrap methods. In order to derive the bootstrap invariance principle, we need the following assumption on the tuning parameter  $\ell$ .

**Assumption 2.**  $\ell = o(T^{1/2})$  and  $\ell \rightarrow \infty$  as  $T \rightarrow \infty$ .

This assumption is comparable to the block length assumption in Paparoditis and Politis (2003) and Palm et al. (2011), as well as to the bandwidth assumption in De Jong and Davidson (2000) needed for processes with stochastic trends (Assumption 4(ii)).

The bootstrap invariance principle is now given in the following theorem.

**Theorem 1.** *Let the conditions of Lemma 1 and Assumption 2 hold. Let the bootstrap sample  $\{u_t^*\}$  be generated according to (3), where  $\{\xi_t^*\}$  is generated using the BWB, DWB or AWB method. Then as  $T \rightarrow \infty$ ,*

$$M_T^*(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} u_t^* \xrightarrow{d^*}_p M(r). \quad (4)$$

The proof of the bootstrap invariance principle evolves around two major steps. First, it needs to be shown that, conditionally on the sample,  $M_T^*(r)$  can be written as a Gaussian process with independent increments. While this follows directly for the BWB, in the Appendix it is shown that for the DWB and AWB one can derive that

$$M_T^*(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} z_t^* \tilde{U}_{t,r}, \quad \text{where } \tilde{U}_{t,r} = \sum_{s=t}^{\lfloor Tr \rfloor} c_{s,t,\ell} u_s, \quad (5)$$

and  $\{z_t^*\}_{t=1}^T$  are independent standard normal variables. The specific form of the non-random weights  $c_{s,t,\ell}$  depends on the method used. For example, for the AWB we have  $c_{s,t,\ell} = \theta^{(s-t)/\ell}$ , so that a larger value for  $\ell$  will ensure that  $\tilde{U}_{t,r}$  becomes smoother. While this is beneficial for capturing serial dependence, it is not for capturing abrupt changes in the volatility, and therefore the choice of  $\ell$  is a trade off as discussed in Remark 6.

The second major step in the proof shows that the MWB methods correctly replicate both the diagonal and off-diagonal elements of the long-run variance-covariance matrix over

all fractions of the sample. Define the long-run covariance matrix of a fraction  $r$  of the data as

$$\Omega(r) = \lim_{T \rightarrow \infty} \Omega_T(r) = \int_0^r \sigma(s) \Lambda \sigma(s)' ds, \quad (6)$$

where

$$\Omega_T(r) = T^{-1} \mathbb{E} \left( \sum_{t=1}^{\lfloor Tr \rfloor} u_t \right) \left( \sum_{t=1}^{\lfloor Tr \rfloor} u_t \right)'. \quad (7)$$

The bootstrap variance process can then be expressed as

$$\Omega_T^*(r) = \mathbb{E}^* M_T^*(r) M_T^*(r)' = T^{-1} \sum_{s=1}^{\lfloor Tr \rfloor} \sum_{t=1}^{\lfloor Tr \rfloor} u_s u_t \mathbb{E}^* \xi_s^* \xi_t^*. \quad (8)$$

The duality between variance estimation and the modified wild bootstrap methods can then be exploited by recognizing (8) as a kernel variance estimator for a fraction  $r$  of the sample, such that results on such kernel variance estimators can be used to show that  $\Omega_T^*(r)$  consistently estimates  $\Omega(r)$  uniformly for  $r \in [0, 1]$ .

### 3 Unit Root Testing

To illustrate the potential of the derived bootstrap invariance principle, in this section we apply it to multivariate unit root testing. A univariate bootstrap unit root test simply follows as a special case by taking  $N = 1$  everywhere. As such we will not discuss this specific case further but rather include it implicitly in our multivariate analysis.

Let the series  $y_t = (y_{1,t}, \dots, y_{N,t})'$  be generated as

$$y_t = \beta' z_t + x_t, \quad t = 1, \dots, T, \quad (9a)$$

$$x_t = \rho x_{t-1} + u_t, \quad x_0 = 0, \quad (9b)$$

where  $u_t$  is the same as in (1),  $z_t = (1, t, \dots, t^{\kappa})'$ ,  $\beta = (\beta_1, \dots, \beta_N)$ , where  $\{\beta_i\}_{i=1}^N$  are  $1 + \kappa \times 1$  coefficient vectors, and  $\rho = \text{diag}(\rho_1, \dots, \rho_N)$ , where  $\rho_1, \dots, \rho_N$  are scalar autoregressive parameters. If  $\rho_i = 1$ , the  $i$ th variable  $y_{i,t}$  has a unit root, whereas it is  $I(0)$  if  $|\rho_i| < 1$ . We consider both fixed alternatives, in which case  $\rho_i$  is independent of  $T$ , and local alternatives, in which case  $\rho = \rho_T$ , and  $\rho_T = 1 - c/T$ , where  $c = \text{diag}(c_1, \dots, c_N)$ . The assumption that the initial values  $x_0$  are equal to zero is made for simplicity only and can easily be dispensed with.

A typical application of unit root testing in a multivariate framework is found in panel unit root testing, where the goal is to test the panel null hypothesis that all variables, or

cross-section units, have a unit root versus the alternative that there are some cross-section units (not necessarily all) that are  $I(0)$ . Clearly the qualification “some” is rather vague and can be interpreted in different ways. This is done on purpose here as different alternative hypotheses are appropriate for different applications. Formally we work here with the alternative hypothesis that a least one unit is stationary, so we test  $H_0 : \rho_i = 1$  for all  $i = 1, \dots, N$  vs.  $H_1 : \rho_i < 1$  for some  $i = 1, \dots, N$ , though our framework easily lends itself to adaptation of different hypotheses. Pesaran (2012) provides a more extensive discussion of the interpretation of panel unit root testing.

There is a large range of potential test statistics that one can employ to test these hypotheses. Each specific test statistic has its own quirks in the development of its theoretical properties. This may concern specific calculations required for the type of detrending performed to accommodate the proofs for parametric or non-parametric correction aimed at accounting for serial correlation. To avoid losing the focus of our main investigation of the bootstrap methods in a quagmire of unrelated technicalities induced by the choice of test statistic, in the theoretical development we focus for illustrative purposes on a very simple test statistic, namely the average, or group-mean, of ordinary least squares (OLS) detrended Dickey-Fuller (DF) coefficient-tests:

$$\tau_{gm} = \frac{T}{N} \sum_{i=1}^N \frac{\sum_{t=2}^T y_{i,t-1}^d \Delta y_{i,t}^d}{\sum_{t=2}^T (y_{i,t-1}^d)^2}, \quad (10)$$

where  $y_t^d = y_t - \hat{\beta}' z_t$  and  $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_N)$ , where  $\{\hat{\beta}_i\}_{i=1}^N$  is the OLS estimator of  $\beta_i$  in a regression of  $y_{i,t}$  on  $z_t$ . As mentioned above there are numerous ways in which this test can be modified to suit one’s needs. For example, a different detrending method could be used. Also, the test statistic could be modified to account for serial correlation by adding lag augmentation or a nonparametric correction. The individual statistics could be combined in a different way in the panel dimension as well, such as by considering a pooled test or the median of the individual unit root test statistics (cf. Palm et al., 2011).<sup>4</sup> Given that none of these issues fundamentally change the setup nor the way the bootstrap should be implemented, we leave all such issues aside and focus on the simple statistic in (10).

We next describe the general bootstrap algorithm to perform this test. Again, variations in specific steps of the algorithm are possible; for example, in Step 1 the method of detrending as well as the estimation method for  $\rho$  can vary. As long as these methods are chosen properly, these do not affect any of our asymptotic results. Therefore we do not go into detail on the different possibilities one has here. Methods of detrending are discussed in detail in Smeeke (2013), while different ways to estimate  $\rho$  are discussed in Paparoditis and Politis (2003,

---

<sup>4</sup>As discussed in Di Iorio and Fachin (2009), the type of panel test should depend on the form of the alternative hypothesis. Smeeke (2011) expands on this concepts by considering a sequential bootstrap procedure based on the theory in Palm et al. (2011) to determine the number of stationary units, also see Hanck (2009) and Moon and Perron (2012).

Remark 2.3).

**Algorithm 1.**

1. Let  $y_t^d = y_t - \hat{\beta}' z_t$  for  $t = 1, \dots, T$ . Let  $\hat{\rho} = \text{diag}(\hat{\rho}_1, \dots, \hat{\rho}_N)$ , where

$$\hat{\rho}_i = \frac{\sum_{t=2}^T y_{i,t-1}^d \Delta y_{i,t}^d}{\sum_{t=2}^T (y_{i,t-1}^d)^2}, \quad \text{for } i = 1, \dots, N,$$

and obtain residuals  $\{\hat{u}_t\}_{t=1}^T$  where  $\hat{u}_1 = y_1^d$  and

$$\hat{u}_t = y_t^d - \hat{\rho} y_{t-1}^d, \quad \text{for } t = 2, \dots, T.$$

2. Generate scalar random variables  $\xi_1^*, \dots, \xi_T^*$  according to an MWB method and let

$$u_t^* = \xi_t^* \hat{u}_t.$$

3. Construct the bootstrap sample as  $x_t^* = x_{t-1}^* + u_t^*$  with  $x_0^* = 0$  and let  $y_t^* = x_t^*$ .
4. For  $i = 1, \dots, N$ , calculate  $y_{i,t}^{*d} = y_{i,t}^* - \hat{\beta}_i^{*'} z_t$ , where  $\hat{\beta}_i^*$  is found by OLS estimation in a regression of  $y_{i,t}^*$  on  $z_t$ , and obtain

$$\tau_{gm}^* = \frac{T}{N} \sum_{i=1}^N \frac{\sum_{t=2}^T y_{i,t-1}^{*d} \Delta y_{i,t}^{*d}}{\sum_{t=2}^T (y_{i,t-1}^{*d})^2}. \quad (11)$$

5. Repeat steps 2 to 4  $B$  times and select the bootstrap critical value, say  $c_{\alpha,B}^*$ , as the  $\alpha$ -quantile of the ordered bootstrap statistics. Reject  $H_0$  if  $\tau_{gm} < c_{\alpha,B}^*$ .

To allow for a general form of detrending, some further notation is needed. Let  $\delta_T = \text{diag}(1, T^{-1}, \dots, T^{-\kappa})$  and  $Z(r) = (1, r, \dots, r^\kappa)'$ , such that  $\delta_T z_{[Tr]} \rightarrow Z(r)$  for all  $r \in [0, 1]$ . Then for an  $N$ -variate continuous martingale  $G(\cdot)$  on  $[0, 1]$ , let  $G^d(r) = G(r) - \int_0^1 G(s) Z(s)' ds \left( \int_0^1 Z(s) Z(s)' ds \right)^{-1} Z(r)$  be its “detrended” version. We can now state the following theorem.

**Theorem 2.** *Let  $y_t$  be generated by (9) and let Assumptions 1 and 2 hold. Let  $c_\alpha^* = \text{plim}_{B \rightarrow \infty} c_{\alpha,B}^*$ , where  $c_{\alpha,B}^*$  is defined in step 5 of Algorithm 1 and obtained using either the BWB, DWB or AWB method. Then as  $T \rightarrow \infty$ , we have the following results:*

- (a) *If  $\rho = \rho_T = 1 - c/T$  with  $c \geq 0$ ,*

$$\mathbb{P}(\tau_{gm} < c_\alpha^*) \rightarrow \mathbb{P}(Y_c < F_0^{-1}(\alpha)), \quad \text{for any } 0 \leq \alpha \leq 1, \quad (12)$$

where  $F_c^{-1}(\alpha)$  is the  $\alpha$ -quantile of the random variable

$$Y_c \sim \frac{1}{N} \sum_{i=1}^N \frac{M_{c,i}^d(1)^2 - M_{c,i}^d(0)^2 - \bar{\sigma}_{i,i}}{2 \int_0^1 M_{c,i}^d(r)^2 dr}, \quad (13)$$

where  $M_{c,i}^d(r)$  is the  $i$ -th element of  $M_c^d(r)$ ,  $M_c(r) = \int_0^r e^{-(r-s)c} dM(s)$ , and  $\bar{\sigma}_{i,i}$  is the  $(i, i)$ -th element of  $\bar{\Sigma} = \int_0^1 \Sigma(s) ds$ .

(b) If  $\rho < 1$  and  $\rho$  does not depend on  $T$ ,

$$\mathbb{P}(\tau_{gm} < c_\alpha^*) \rightarrow 1.$$

Part (a) establishes the asymptotic local power function of the bootstrap test. A particular consequence of part (a) of Theorem 2 is that under  $H_0$ , where  $c = 0$ , it follows that

$$\mathbb{P}(\tau_{gm} < c_\alpha^*) \rightarrow \alpha, \quad \text{for any } 0 \leq \alpha \leq 1, \quad (14)$$

and thus the bootstrap unit root test has the correct asymptotic size. Moreover, it follows from part (a) that the bootstrap test has the same local power function as a size-corrected asymptotic test, which is the best one can do for a given test statistic; that is, the local power function cannot be improved while retaining the same test statistic. Part (b) considers fixed alternatives and establishes that the test is consistent.

**Remark 8.** When one considers the asymptotic distribution of the test statistic given in (13), it becomes clear that it is a functional of  $M(\cdot)$  obtained in the invariance principle. As such, the limit distribution contains nuisance parameters stemming from both the long-run covariance matrix  $\Lambda$  and the volatility process  $\sigma(\cdot)$ . While many techniques exist to estimate or correct for the elements of  $\Lambda$ , they are generally only applicable if  $N$  is rather small compared for  $T$ . Even in typical macroeconomic panel data models, where  $N$  may be smaller than  $T$ , but not much smaller, estimating the off-diagonal covariance elements is rather difficult (though recently some methods have been explored for such panel data, see e.g. Pedroni, Vogelsang, Wagner, and Westerlund, 2008, and Breitung and Cubadda, 2011). Second, the limit distribution depends in a complicated way on the volatility process. Successful correction for these parameters is complicated even in the univariate case. Bootstrap methods such as the MWB that are able to successfully replicate both sources of nuisance, therefore remove the need for corrections asymptotically.<sup>5</sup>

---

<sup>5</sup>In small samples it may still be beneficial to attempt to correct for the nuisance parameter. As is well known from the bootstrap literature on asymptotic refinements, the bootstrap performs best on asymptotically pivotal statistics. Also see Palm, Smeekes, and Urbain (2008) for Monte Carlo evidence that confirms this in the context of univariate unit root testing.



**Remark 9.** As mentioned in Remark 1, cointegration between units is allowed by allowing a reduced rank of  $\Lambda$ . The specification considered here with  $\rho$  a diagonal matrix might appear to be more restrictive than the usual vector error correction model (VECM) used for cointegration analysis (see e.g. Johansen, 1995), but this is not so. Smeekees and Urbain (2014) start from a typical VECM representation with linear process errors

$$\Delta x_t = \alpha \beta' x_{t-1} + \Psi(L) \varepsilon_t$$

and show that one may equivalently write  $\Delta x_t = a(L)^{-1} B(L) \varepsilon_t$ , where  $a(z)$  and  $B(z)$  are lag polynomials defined in their paper. They also show that it follows from the common trends representation (Johansen, 1995, Theorem 4.2) that the long-run covariance matrix of  $\Delta x_t$  has reduced rank if cointegration is present. Hence, the usual VECM representation yields a very similar representation to the one we use, the only (non-fundamental) differences being the addition of non-stationary volatility here and mixing instead of linear process errors.

**Remark 10.** While the MWB methods are theoretically valid in the presence of cointegration, this result should be regarded with caution. If cointegration between units is present,  $\Lambda$  and therefore  $\Omega(r)$  are of reduced rank. However, as we do not impose the rank restriction in the bootstrap, in finite samples the bootstrap data will have a full rank long-run covariance matrix with probability 1. Therefore, we can still expect considerable size distortions in finite samples if the units are cointegrated, which is confirmed by our simulation study. We therefore do not put major emphasis on the cointegration case in this paper. Palm et al. (2011) provide an in-depth analysis of cross-unit cointegration which is appropriate for the current paper as well.

**Remark 11.** The asymptotic validity of the SWB in a univariate framework is well known (cf. Cavaliere and Taylor, 2009a,b; Smeekees and Taylor, 2012). However, in the multivariate context considered here, the SWB when applied equation by equation, is not asymptotically valid. Smeekees and Urbain (2014) prove the invalidity of the autoregressive sieve bootstrap with i.i.d. resampling in a homoskedastic setting. They show that the cause of the invalidity is the inability of the i.i.d. bootstrap to capture serial correlation across units, which is unrelated to the treatment of heteroskedasticity. As the wild bootstrap only differs from the i.i.d. bootstrap in its treatment of heteroskedasticity, and suffers from the same inability to capture these dynamic correlations, their invalidity result directly extends to the SWB. If the dependence between units is of contemporaneous nature only, the SWB is valid, but if the dependence is of a dynamic nature as well, then asymptotic validity of the SWB is not guaranteed. In particular, while there might be certain specific parameter combinations for which the invalidity is not very severe, for other parameter combinations the SWB may fail seriously. Smeekees and Urbain (2014) illustrate this by considering a simple DGP for which analytical expressions of the asymptotic size can be derived, and demonstrate that in this

model the extent of the asymptotic size violations depends crucially on the choices of the parameter values in the model.

We note that in contrast to the equation-by-equation approach discussed above, the VAR specification of the SWB is valid for these models (cf. Cavaliere et al., 2010). However, for the typical applications that are considered in the nonstationary panel data literature, the dimension of the system is too large to estimate a full VAR model. The MWB methods therefore provide an attractive and valid alternative in this multivariate framework.

**Remark 12.** A different type of application can be found in cointegration analysis (cf. Johansen, 1995). We do not consider this here as the dimensions of the system (here denoted by  $N$ ) in which a full-blown cointegration analysis is possible are fairly small; typically no much larger than 5. We believe the MWB methods have an added value in systems where  $N$  is larger, so a cointegration analysis is not the most appropriate framework here. Obviously though, the MWB methods can be applied in that framework as well.

**Remark 13.** In the literature often a distinction is made between multivariate time series application with finite  $N$  and “true” panel applications with  $N$  increasing to infinity (usually jointly with  $T$ ). According to this classification our method could not be classified as a panel data method. In practice, however, the distinction between multivariate time series and panel data is not clear cut, as there is no rule of thumb that describes what kind of asymptotics is most appropriate for which combinations of, in practical applications inherently finite,  $N$  and  $T$ . In particular, there is a large grey area, where both types of asymptotic frameworks may be applicable, between the typical multivariate time series applications such as described in Remark 12 and “proper” panel data studies where  $N$  is of the same order as  $T$ . As such, in the recent nonstationary panel data literature quite a few contributions have adopted an asymptotic framework with finite  $N$ , in particular when using resampling methods (cf. Palm et al., 2011, p. 89). In our simulation study in Section 4.2 we investigate the applicability of our methods to different combinations of  $N$  and  $T$ .

## 4 Monte Carlo Simulations

In this section we analyze the performance of the MWB in finite samples in a Monte Carlo simulation study. As in the previous section we focus here on unit root testing. We consider both univariate unit root testing and multivariate testing. The univariate setting serves as a reference point for the MWB methods in comparison with well established alternatives, in particular the sieve wild bootstrap. The major application of the methods however lies in multivariate applications, which is investigated in a second simulation study involving panel unit root testing.

## 4.1 Univariate Unit Root Tests

We start by analyzing the performance of the methods in univariate time series. It is well known from simulation studies that the SWB performs extremely well in univariate time series. As such it provides a useful benchmark for the MWB methods. Our aim in this Monte Carlo study is therefore to examine how close the MWB methods get to the benchmark.

We consider the DGP

$$\begin{aligned} y_t &= \rho_T y_{t-1} + u_t, & u_t &= \phi u_{t-1} + v_t + \theta v_{t-1}, \\ v_t &= \sigma_t \varepsilon_t, & \varepsilon_t &\sim N(0, 1), \end{aligned}$$

where we take  $\rho_T = 1 - c/T$  with  $c = 0$  to analyze size and  $c = 20$  for power. Sample sizes  $T = 50, 100, 250$  are considered. All results are based on 1000 simulations and 199 bootstrap replications and were obtained using GAUSS 12.<sup>6</sup>

We use the augmented Dieckmann-Fuller (ADF)  $t$ -test with OLS demeaning, where in the bootstrap ADF tests lag lengths  $p$  and  $p^*$  are selected using the rescaled modified Akaike information criterion (RSMAIC) developed by Cavaliere, Phillips, Smeekes, and Taylor (2012), with a maximum lag length of  $\lfloor 12(T/100)^{1/4} \rfloor$ . We apply the four bootstrap methods DWB, BWB, AWB and SWB to this test, and we also add the moving block bootstrap (MBB) test of Paparoditis and Politis (2003) which is only valid under homoskedasticity. For the MBB, DWB and BWB we take  $\ell = 1.75T^{1/3}$ , while for the AWB we take  $\gamma = \theta^\ell$  with  $\theta = 0.01$ . This value for  $\ell$  was considered in Palm et al. (2011) to lead to satisfactory results for their analysis of the moving block bootstrap in multivariate unit root testing. As block length selection methods remain largely undeveloped for nonstationary time series, we follow their lead and choose this fixed function of the sample size. For the SWB we select  $q$  by RSMAIC, with a maximum lag length of  $\lfloor 12(T/100)^{1/4} \rfloor$ . Obviously the data dependent specification of the lag length in the sieve bootstrap gives it an advantage compared with the other methods; however it also has this advantage in practical applications given the difficulties of selecting block lengths in a data dependent way. We therefore do not believe the comparison to be unfair, though of course it should be kept in mind when evaluating the results.

For the volatility we consider the following four specifications:

- Homoskedasticity [HOM]:  $\sigma_t = 1$  for all  $t = 1, \dots, T$ .
- Single break in volatility:  $\sigma_t^2 = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)I(\lfloor \tau T \rfloor < t)$ . We set  $\sigma_0 = 1$ ,  $\sigma_1 = 3$  and  $\tau = 0.8$  to achieve a *late positive break* [LPB].
- Smooth transition:  $\sigma_t^2 = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mathbb{S}_t$ , where  $\mathbb{S}_t = (1 + \exp(-\gamma(t - \lfloor \tau T \rfloor)/T))^{-1}$ . We set  $\sigma_0 = 3$ ,  $\sigma_1 = 1$ ,  $\tau = 0.3$  and  $\gamma = 15$  to achieve a *downward smooth transition* [DST].

---

<sup>6</sup>Code to perform the BWB, DWB and AWB unit root tests developed in this paper is available at the website [researchers-sbe.unimaas.nl/stephansmeekes/research/](http://researchers-sbe.unimaas.nl/stephansmeekes/research/).

- Stochastic volatility:  $\sigma_t^2 = \omega^2(t/T)$  where  $\omega^2(s) = \sigma_0^2 \exp(\nu B(s))$  and  $B(s)$  is a standard Brownian motion. We set  $\sigma_0 = 1$  and  $\nu = 4$  to achieve *integrated stochastic volatility* [ISV].

Note that none of the MWB methods have been shown to be valid for the ISV model. As such this specification serves as a robustness check for the MWB methods. Clearly only considering one specification of each of the volatility models is rather restrictive. However, we do not strive to obtain a complete overview with this Monte Carlo study, but rather to gain an initial understanding of how the methods perform. Within each of the three specifications considered above one could consider numerous parameter combinations. Reporting many of those would not result in a clearer picture we believe. Instead we therefore choose just one parameter combination per model that we believe is representative for that heteroskedastic model. Simulation results for a wider set of parameter combinations are available on request.

Table 1 presents the empirical rejection frequencies of the tests for the models described above under homoskedasticity. The top panel reports size of the tests under homoskedasticity. Here, five combinations of different values for  $\phi$  and  $\theta$  are considered, the values of which are given in the table. While all five tests have very reasonable size properties under homoskedasticity, there are still some noticeable differences, in particular for  $T = 50$ . The MBB and BWB tend to be slightly oversized, while the SWB tends to be undersized.

The middle panel presents the corresponding power results for the local alternative  $c = 20$ . For small sample sizes, the power for all tests is quite sensitive to the dynamic specification, although this effect disappears for larger sample sizes. The SWB appears to be somewhat less powerful than the MWB methods, in particular for the smaller sample sizes, which is most likely due to its tendency to be undersized.

Finally, the bottom panel presents size results for the three heteroskedastic models. To save space and to focus on the effect of the volatility process, we only report results for the case where  $\phi = \theta = 0$ , as we already analyzed the effects of serial correlation for the homoskedastic model. Also, we do not report powers for the heteroskedastic models as these are difficult to interpret due to distortions in the size of the tests. As expected, the SWB method has the best size properties under the heteroskedastic models. The MBB method suffers from serious size distortions that do not decrease if the sample size increases. Of course, this is to be expected as the MBB is not valid in the presence of unconditional heteroskedasticity. The MWB methods control size fairly well. They do suffer from size distortions for small sample sizes, but the size distortions clearly become smaller for larger sample sizes, which one would expect from the results in Section 3. The AWB seems to control size best of the MWB methods. For the AWB in particular, but also for the MWB methods in general, the choice of the tuning parameter represents a trade off between controlling for serial correlation and volatility changes. In particular, if  $\ell$  (or  $\theta$  for the AWB) increases, more serial dependence can be captured, but, in attempting to replicate the variance of the partial sum, the volatility

changes get smoothed out, and are thus not mimicked perfectly. Unreported simulations with the AWB clearly show that if  $\theta$  is increased, performance in the models with serial correlation improves while that in heteroskedastic models deteriorates, and vice versa. The choice of  $\theta = 0.01$  represents a compromise between both effects.<sup>7</sup>

Concluding, our results show that the MWB methods are not far off the benchmark set by the SWB. Under homoskedasticity the DWB and AWB even perform slightly better than the SWB. The MBB also performs well under homoskedasticity, but it falters under the heteroskedastic models considered. Under heteroskedasticity the SWB confirms its excellent performance found in earlier literature, though the MWB methods are not far off, and in particular the performance of the AWB is quite satisfactory. Overall little is lost, though also not gained, when using the MWB methods rather than the SWB method for univariate unit root testing. Therefore, to really appreciate the added value of the MWB methods, we now turn to multivariate unit root testing, a setting in which the SWB is invalid asymptotically.

## 4.2 Multivariate (Panel) Unit Root Tests

In this section we study the application of the MWB methods to multivariate unit root testing in a Monte Carlo experiment. In particular, we consider an application in the context of panel unit root testing. As explained in Remark 11, the SWB is not valid if dynamic cross-sectional dependence is present in the DGP, although as pointed out by Smeekes and Urbain (2014), the invalidity of the sieve bootstrap is not always detectable in simulation studies. Regardless, one should be very cautious about applying an invalid bootstrap method, in particular as Smeekes and Urbain (2014) show that also in finite samples the oversize can be quite large as the exact values of the parameters in the model determinate the extent of the invalidity of the sieve bootstrap. As such the asymptotically valid MWB methods provide an appealing alternative in this multivariate context.

We assume that  $y_{i,t}$ ,  $i = 1, \dots, N$ ,  $t = 1, \dots, T$ , are generated by the following panel data model:

$$\begin{aligned} y_{i,t} &= x_{i,t} + \lambda_i f_t, & f_t &= \rho_f f_{t-1} + \zeta_t, \\ x_{i,t} &= \rho_i x_{i,t-1} + u_{i,t}, & u_t &= \Phi u_{t-1} + v_t + \Theta v_{t-1}, \\ v_{i,t} &= \sigma_t \varepsilon_{i,t}, & \varepsilon_t, \zeta_t &\sim N(0, 1), \end{aligned} \tag{15}$$

where  $f_t$  is a common factor with loadings  $\lambda_i$ ,  $x_{i,t}$  are idiosyncratic components, and  $\sigma_t$  controls the potential heteroskedasticity. Sample sizes considered are combinations of  $T = 25, 50, 100$  and  $N = 10, 25$ . The combination of the parameters  $\rho_1, \dots, \rho_N$  and  $\rho_f$  determine if the cross-section units have a unit root (see p. 86 of Palm et al., 2011, for more details). We test the null hypothesis that all units have a unit root against the alternative that at

---

<sup>7</sup>As such, any successful method for selecting the tuning parameters should also take the shape of the volatility process into account.

Table 1: Empirical rejection frequencies of univariate bootstrap ADF tests

VP	$c$	$T$	$\phi$	$\theta$	MBB	BWB	DWB	AWB	SWB
HOM	0	50	0	0	0.054	0.068	0.042	0.046	0.038
			0.5	0	0.057	0.044	0.041	0.053	0.030
			-0.5	0	0.062	0.061	0.047	0.043	0.038
			0	0.5	0.061	0.053	0.045	0.060	0.028
			0	-0.5	0.083	0.077	0.062	0.062	0.058
		100	0	0	0.066	0.060	0.051	0.061	0.050
			0.5	0	0.060	0.060	0.055	0.063	0.042
			-0.5	0	0.062	0.057	0.051	0.038	0.042
			0	0.5	0.067	0.068	0.060	0.065	0.042
			0	-0.5	0.072	0.075	0.055	0.049	0.061
		250	0	0	0.066	0.075	0.070	0.073	0.059
			0.5	0	0.061	0.059	0.057	0.061	0.047
			-0.5	0	0.065	0.057	0.057	0.050	0.046
			0	0.5	0.060	0.063	0.058	0.053	0.044
			0	-0.5	0.065	0.067	0.051	0.059	0.056
HOM	20	50	0	0	0.614	0.543	0.522	0.555	0.544
			0.5	0	0.132	0.156	0.128	0.146	0.101
			-0.5	0	0.548	0.471	0.468	0.470	0.452
			0	0.5	0.350	0.332	0.301	0.341	0.273
			0	-0.5	0.598	0.575	0.554	0.569	0.542
		100	0	0	0.633	0.568	0.568	0.589	0.593
			0.5	0	0.337	0.315	0.329	0.391	0.324
			-0.5	0	0.534	0.456	0.469	0.466	0.461
			0	0.5	0.393	0.362	0.361	0.416	0.364
			0	-0.5	0.497	0.477	0.440	0.462	0.439
		250	0	0	0.713	0.688	0.661	0.704	0.689
			0.5	0	0.608	0.574	0.563	0.588	0.588
			-0.5	0	0.678	0.632	0.637	0.661	0.638
			0	0.5	0.569	0.536	0.541	0.531	0.531
			0	-0.5	0.586	0.559	0.541	0.573	0.534
LPB	0	50	0	0	0.030	0.048	0.042	0.046	0.046
		100	0	0	0.023	0.045	0.048	0.055	0.050
		250	0	0	0.026	0.051	0.050	0.047	0.047
DST		50	0	0	0.276	0.147	0.095	0.097	0.075
		100	0	0	0.268	0.134	0.082	0.084	0.057
		250	0	0	0.251	0.096	0.078	0.075	0.069
ISV		50	0	0	0.153	0.095	0.070	0.074	0.064
		100	0	0	0.165	0.086	0.074	0.061	0.055
		250	0	0	0.149	0.063	0.061	0.051	0.053

Notes: ‘VP’ refers the volatility process used to generate  $\sigma_t$ .

least one unit is  $I(0)$ . To analyze size we either need  $\rho_i = 1$  for all  $i = 1, \dots, N$  or  $\rho_f = 1$  in combination with  $\lambda_i \neq 0$  for all  $i = 1, \dots, N$ . For the power analysis we work under the alternative hypothesis by letting  $\rho_i \sim U[0.8, 1]$  for  $i = 1, \dots, N$  and  $\rho_f = 0.95$ . For the volatility process  $\sigma_t$  we consider the same specifications as in the univariate simulation study. Note that for simplicity we let the volatility process affect all units in an identical way.

In the panel data literature, specific attention is paid to dependence across units, or to remain in panel data terminology, cross-sectional dependence. This can be of purely contemporaneous nature but it can also be dynamic, for example of autoregressive kind where lags of one unit affect the present of another units. The DGP in (15) is general and flexible enough to generate all such forms of dependence. To keep the analysis manageable, we focus on five intuitive dependence structures (DS):

- DS IID: there is no serial or cross-sectional correlation. Specifically, we take  $\lambda_i = 0$  for all  $i = 1, \dots, N$  and  $\Phi = \Theta = 0$ .
- DS VAR1: a “spatial”-type VARMA structure, where the highest dependence is on closest neighbors. This generates both serial and dynamic cross-sectional dependence. Specifically, we let  $\lambda_i = 0$  for all  $i = 1, \dots, N$  and we generate  $\Phi$  and  $\Theta$  as  $\Xi$  and  $\Omega$  defined in Palm et al. (2011, p. 92).
- DS VAR2: there is one dominant unit that affects all units dynamically. Specifically, we let  $\lambda_i = 0$  for all  $i = 1, \dots, N$ , while we set  $\Phi = (\phi_{i,j})_{i,j=1}^N$  with

$$\phi_{i,j} = \begin{cases} \xi_i & \text{if } i = j, \\ \eta_j & \text{if } i = 1 \text{ and } j > 1, \\ 0 & \text{elsewhere,} \end{cases}$$

where  $\xi_1, \dots, \xi_N \sim U[-0.5, 0.5]$  and  $\eta_2, \dots, \eta_N \sim U[-0.8, 0.8]$ . Furthermore we let  $\Theta = \Omega$  as for model VAR1.

- DS CF: there is one common factor generating cross-sectional dependence. Specifically, we take  $\lambda_i \sim U[-1, 3]$  for  $i = 1, \dots, N$  and  $\Phi = \Theta = 0$ .
- DS CC: there is an integrated common factor, but the idiosyncratic components are stationary. This setup implies that units are cross-sectional cointegrated, where the cointegration vectors are determined by their factor loadings. Specifically, we take  $\lambda_i \sim U[-1, 3]$  and  $\rho_i \sim U[0.8, 1]$  for  $i = 1, \dots, N$  while setting  $\Phi = \Theta = 0$ .

DS IID is the most basic panel data setup. VAR1 and VAR2 represent two plausible models for situations in which dynamic cross-sectional dependence is present, and for which the SWB is asymptotically invalid. CF and CC are typical workhorse structures in the panel unit root literature when cross-sectional dependence is of interest. Structure CF in particular

is the standard model in the panel data literature for cross-sectional dependence. While it does not have any particular significance in our analysis, it is considered here as well in order to remain consistent with the existing panel unit root literature. DS CC is, like CF, a setting that is often considered in simulation studies for panel unit root tests, as typically most panel unit root tests fail for this case, either asymptotically or in finite samples). The multivariate MBB test of Palm et al. (2011) is valid for CC (as well as for all four other models) under homoskedasticity. However, as most other panel unit root tests, it still suffers from small sample size distortions in this setting. While the MWB tests are asymptotically valid for CC, we expect them to suffer from finite sample size distortions, for the same reason as the MBB suffers; that is, while asymptotically the reduced rank of  $\Lambda$  can be replicated by the bootstrap methods, in finite samples the long-run bootstrap covariance matrix will always be of full rank (see Section 4.2 of Palm et al., 2011, for more details). As the cross-cointegration setting is very peculiar to the panel unit root application, we believe that in general not too much weight should be given to the performance of the tests in this case. However, to remain consistent with the panel unit root literature, we do consider it here as well.

As test statistic we consider the average of ADF  $t$ -statistics with OLS demeaning, which is the panel counterpart of the unit root test considered in Section 4.1. Again we select lag lengths for the (bootstrap) ADF tests by RSMAIC. We consider the same bootstrap methods as in the previous simulation study, where the considered multivariate SWB is essentially the bootstrap procedure proposed in Smeekes and Urbain (2014) with the i.i.d. bootstrap in step 3 replaced by the wild bootstrap. The multivariate MBB unit root test was considered by Palm et al. (2011) and shown to be asymptotically valid for homoskedastic models. However, just as its univariate counterpart, it is not valid in the presence of unconditional heteroskedasticity. The tuning parameters for the bootstrap methods are selected in the same way as in Section 4.1. Our results are again based on 1000 simulations and 199 bootstrap replications.

Table 2 presents the empirical sizes under homoskedasticity for the five structures described above. In general all methods suffer from undersize, though it is particularly severe for the BWB and DWB methods. The undersize of the tests is exacerbated when  $N$  increases, but is diminished for increasing  $T$ . As the proof of asymptotic validity only assumes  $T$  increases to infinity, these results are not at odds with the theoretical results. It is clear that all methods work best if  $T$  is clearly larger than  $N$ . Still, if one considers the case where  $T = 100$  and  $N = 25$ , empirical sizes for AWB and SWB in particular are acceptable for much larger values of  $N$  than those that allow a full-blown multivariate time series analysis such as discussed in Remark 12.

Three further features are noticeable. First, the MBB does not suffer from undersize but is considerably oversized instead. Remarkably, if one uses the ADF coefficient test rather than the  $t$ -test such as Palm et al. (2011) do, the MBB has very reasonable size properties, comparable with the AWB and SWB methods. We do not know what causes the MBB



Table 2: Empirical rejection frequencies under  $H_0$  of bootstrap group-mean ADF tests under homoskedasticity

VP	N	T	DS	MBB	BWB	DWB	AWB	SWB
HOM	10	25	IID	0.094	0.013	0.003	0.031	0.021
			VAR1	0.126	0.027	0.012	0.062	0.037
			VAR2	0.081	0.009	0.004	0.026	0.024
			CF	0.131	0.042	0.021	0.067	0.054
			CC	0.203	0.077	0.057	0.132	0.120
		50	IID	0.081	0.007	0.004	0.030	0.018
			VAR1	0.075	0.022	0.016	0.049	0.038
			VAR2	0.114	0.040	0.028	0.080	0.061
			CF	0.096	0.055	0.042	0.055	0.037
			CC	0.209	0.132	0.103	0.160	0.142
		100	IID	0.111	0.015	0.021	0.038	0.031
			VAR1	0.098	0.040	0.032	0.068	0.077
			VAR2	0.097	0.029	0.030	0.065	0.051
			CF	0.126	0.064	0.057	0.070	0.053
			CC	0.419	0.244	0.301	0.385	0.361
	25	25	IID	0.075	0.000	0.000	0.011	0.007
			VAR1	0.091	0.002	0.000	0.015	0.008
			VAR2	0.044	0.000	0.000	0.006	0.004
			CF	0.128	0.030	0.009	0.038	0.032
			CC	0.181	0.083	0.043	0.129	0.090
		50	IID	0.102	0.002	0.001	0.015	0.013
			VAR1	0.090	0.003	0.003	0.022	0.020
			VAR2	0.094	0.003	0.002	0.029	0.017
			CF	0.147	0.051	0.036	0.070	0.052
			CC	0.215	0.090	0.072	0.146	0.123
		100	IID	0.114	0.004	0.002	0.022	0.025
			VAR1	0.128	0.006	0.012	0.039	0.049
			VAR2	0.126	0.008	0.017	0.046	0.049
			CF	0.126	0.034	0.029	0.050	0.042
			CC	0.291	0.209	0.210	0.271	0.266

Notes: ‘VP’ refers the volatility process used to generate  $\sigma_t$ ; ‘DS’ refers to the specification used to generate the dependence structure.

method to deteriorate while using the  $t$ -test. Second, except from the undersize when  $N$  is large relative to  $T$ , the SWB has good size properties, even for the models for which it is asymptotically not valid. As mentioned above, Smeeke and Urbain (2014) already found that the invalidity of the sieve bootstrap is not always detectable from simulation studies as the extent of the invalidity of the sieve bootstrap depends on specific parameter combinations. Third, as expected, all tests suffer from oversize for the CC specification.

Table 3 presents the corresponding power results under homoskedasticity. Note that here only the first four models need to be considered: DS CC collapses to the same specification as CF under the alternative. For all tests, the empirical power increases with both  $T$  and  $N$ . Given that we consider fixed alternatives here, consistency requires that power increases with  $T$ . It is reassuring that power also increases with  $N$  despite the increasing size distortions found before. This does not mean that the undersize has no effect on power; the BWB and DWB methods, which suffered from rather severe undersize, have considerably lower power than the AWB and SWB methods. The MBB method has highest power of all, but this is an artifact of its too liberal size properties. The powers of the AWB and SWB methods are similar, though the AWB seems to have slightly higher power on average.

Finally, empirical sizes under heteroskedasticity are presented in Table 4. As in the univariate simulation study, to preserve space and to ensure that the results remain interpretable, we do not report size results for specifications other than IID nor do we report power results. The size results for the IID model show that the invalidity of the MBB becomes much more severe in the multivariate setting, with empirical sizes approaching 1. The tendencies for the MWB methods to display oversize for small  $T$  in the univariate setting are countered by the undersize present for increasing  $N$ . This is particularly noticeable for the BWB and DWB. The AWB maintains quite good size properties, and even shows smaller undersize than the SWB.

The overall conclusion of the univariate simulation study that the SWB and AWB perform best remains true in the multivariate case. In the multivariate case the BWB and DWB methods suffer from severe undersize. The MBB is generally oversized, in particular under heteroskedasticity where its empirical size can approach 1. The AWB method performs reasonably well under both homoskedasticity and heteroskedasticity. The same is observed for the SWB, though this may be an artifact of the particular DGPs used in this simulation study, as it is known to be asymptotically invalid if dynamic dependence between units is present. The AWB method however provides an asymptotically valid alternative with similar finite sample properties.

Table 3: Empirical rejection frequencies under  $H_1$  of bootstrap group-mean ADF tests under homoskedasticity

VP	N	T	DS	MBB	BWB	DWB	AWB	SWB
HOM	10	25	IID	0.205	0.025	0.016	0.082	0.072
			VAR1	0.239	0.059	0.031	0.165	0.148
			VAR2	0.126	0.022	0.009	0.071	0.059
			CF	0.175	0.055	0.035	0.102	0.090
		50	IID	0.573	0.169	0.165	0.409	0.373
			VAR1	0.488	0.190	0.186	0.393	0.340
			VAR2	0.666	0.293	0.317	0.558	0.485
			CF	0.174	0.134	0.114	0.139	0.124
		100	IID	0.942	0.705	0.778	0.901	0.878
			VAR1	0.936	0.700	0.763	0.906	0.909
			VAR2	0.815	0.522	0.592	0.795	0.769
			CF	0.609	0.448	0.451	0.559	0.567
	25	25	IID	0.296	0.005	0.002	0.063	0.053
			VAR1	0.281	0.002	0.001	0.073	0.063
			VAR2	0.129	0.003	0.001	0.028	0.025
			CF	0.217	0.034	0.015	0.109	0.098
		50	IID	0.933	0.206	0.277	0.795	0.767
			VAR1	0.599	0.083	0.080	0.450	0.489
			VAR2	0.487	0.056	0.056	0.325	0.318
			CF	0.256	0.137	0.114	0.196	0.174
		100	IID	1.000	0.979	0.995	1.000	1.000
			VAR1	0.995	0.789	0.874	0.986	0.986
			VAR2	0.993	0.790	0.905	0.989	0.988
			CF	0.395	0.304	0.303	0.352	0.346

Notes: see Table 2.

Table 4: Empirical rejection frequencies under  $H_0$  of bootstrap group-mean ADF tests under heteroskedasticity

VP	N	T	DS	MBB	BWB	DWB	AWB	SWB
LPB	10	25	IID	0.006	0.014	0.002	0.033	0.027
		50		0.005	0.024	0.011	0.047	0.045
		100		0.007	0.032	0.029	0.042	0.035
	25	25		0.001	0.000	0.000	0.016	0.015
		50		0.000	0.008	0.000	0.035	0.031
		100		0.000	0.011	0.012	0.041	0.046
	DST	25		0.859	0.083	0.025	0.080	0.051
		50		0.889	0.059	0.029	0.067	0.037
		100		0.891	0.047	0.040	0.064	0.044
ISV	25	25		0.974	0.020	0.003	0.025	0.022
		50		0.993	0.013	0.005	0.032	0.026
		100		0.999	0.014	0.008	0.027	0.027
	10	25		0.302	0.061	0.029	0.052	0.048
		50		0.301	0.051	0.030	0.050	0.033
		100		0.309	0.049	0.049	0.055	0.052
	25	25		0.341	0.037	0.014	0.037	0.022
		50		0.345	0.034	0.018	0.038	0.026
		100		0.347	0.030	0.024	0.039	0.038

Notes: see Table 2.

## 5 Conclusion

In this paper we developed a multivariate invariance principle for three modifications of the wild bootstrap that are designed to capture serial dependence as well as unconditional heteroskedasticity. The bootstrap invariance principle was subsequently applied to multivariate unit root testing and utilized to establish the asymptotic validity of unit root tests based on the modified wild bootstrap methods. The modified wild bootstrap methods have as major advantage over the autoregressive sieve wild bootstrap that it can also be applied if the dimension of the system is relatively large, a setting in which the sieve wild bootstrap is inappropriate. Moreover, the modified wild bootstrap methods do not require the specification of either the form of time-variation in the volatility or the form of the serial dependence, thereby making it a convenient tool for practitioners.

We examined the finite sample performance of the MWB methods in two simulation studies involving univariate and multivariate unit root testing. For univariate unit root testing the MWB methods performed very similar to the SWB, which is known for its excellent performance in small samples. In the multivariate setting, where the SWB is invalid, the MWB methods performed reasonably well, in particular the AWB method. Despite its asymptotic invalidity the SWB did well in the simulations, yet it should not be routinely applied as the good performance may be specific to our simulation DGP. Our simulation studies also clearly

show that the moving block bootstrap is not appropriate in the presence of unconditional volatility changes and can lead to massive size distortions.

The choice of the tuning parameter in the MWB represents a trade off between capturing serial dependence and (abrupt) changes in the volatility. Its selection in practice remains an open issue and is an important topic for future research, even though our simulation study showed that an ad hoc specification of the tuning parameter can lead to satisfactory results in finite samples, in particular for the AWB. Overall, it appears that the MWB methods have much potential for dealing with serial dependence and unconditional heteroskedasticity in a flexible and nonparametric manner.

## References

- Blanchard, O. and J. Simon (2001). The long and large decline in U.S. output volatility. *Brookings Papers on Economic Activity* 32, 135–174.
- Breitung, J. and G. Cubadda (2011). Testing for cointegration in high-dimensional systems. Working paper.
- Canova, F. and M. Ciccarelli (2013). Panel vector autoregressive models: a survey. In L. Kilian, A. Murphy, and T. B. Fomby (Eds.), *VAR Models in Macroeconomics, Financial Econometrics, and Forecasting - New Developments and Applications*, Volume 31 of *Advances in Econometrics*, pp. forthcoming. Emerald Group Publishing Limited.
- Cavaliere, G. (2005). Unit root tests under time-varying variances. *Econometric Reviews* 23, 259–292.
- Cavaliere, G., P. C. B. Phillips, S. Smeeke, and A. M. R. Taylor (2012). Lag length selection for unit root tests in the presence of nonstationary volatility. Cowles Foundation Discussion Papers 1844, Cowles Foundation for Research in Economics, Yale University.
- Cavaliere, G., A. Rahbek, and A. M. R. Taylor (2010). Testing for co-integration in vector autoregressions with non-stationary volatility. *Journal of Econometrics* 158, 7–24.
- Cavaliere, G. and A. M. R. Taylor (2008). Bootstrap unit root tests for time series with nonstationary volatility. *Econometric Theory* 24, 43–71.
- Cavaliere, G. and A. M. R. Taylor (2009a). Bootstrap  $M$  unit root tests. *Econometric Reviews* 28, 393–421.
- Cavaliere, G. and A. M. R. Taylor (2009b). Heteroskedastic time series with a unit root. *Econometric Theory* 25, 1228–1276.
- Davidson, J. (2002). *Stochastic Limit Theory* (2nd ed.). Oxford University Press.

- Davidson, R. and E. Flachaire (2008). The wild bootstrap, tamed at last. *Journal of Econometrics* 146, 162–169.
- De Jong, R. M. and J. Davidson (2000). Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. *Econometrica* 68, 407–423.
- Demetrescu, M. and C. Hanck (2012). Unit root testing in heteroskedastic panels using the Cauchy estimator. *Journal of Business and Economic Statistics* 30, 256–264.
- Di Iorio, F. and S. Fachin (2009). A residual-based bootstrap test for panel cointegration. *Economics Bulletin* 29, 3222–3232.
- Doukhan, P., G. Lang, A. Leucht, and M. H. Neumann (2014). Dependent wild bootstrap for the empirical process. Department of Economics Working Papers Series 14-01, University of Mannheim.
- Francq, C. and J.-M. Zakoïan (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.
- Giné, E. and J. Zinn (1990). Bootstrapping general empirical measures. *Annals of Probability* 18, 851–869.
- Gonçalves, S. and L. Kilian (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics* 123, 89–120.
- Gonçalves, S. and L. Kilian (2007). Asymptotic and bootstrap inference for  $AR(\infty)$  processes with conditional heteroskedasticity. *Econometric Reviews* 26, 609–641.
- Hanck, C. (2009). For which countries did PPP hold? A multiple testing approach. *Empirical Economics* 37, 93–103.
- Hansen, B. E. (1991). Strong laws for dependent heterogeneous processes. *Econometric Theory* 7, 213–221.
- Hansen, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 498–500.
- Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica* 68, 575–603.
- Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.
- Kim, C.-J. and C. R. Nelson (1999). Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle. *Review of Economics and Statistics* 81, 608–616.

- Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. Springer-Verlag.
- Leucht, A. and M. H. Neumann (2013). Dependent wild bootstrap for degenerate  $U$ - and  $V$ -statistics. *Journal of Multivariate Analysis* 117, 257–280.
- Ling, S. and M. McAleer (2002). Stationarity and the existence of moments of a family of GARCH processes. *Journal of Econometrics* 106, 109–117.
- McConnell, M. M. and G. Perez Quiros (2000). Output fluctuations in the United States: what has changed since the early 1980s? *American Economic Review* 90, 1464–1476.
- Moon, H. R. and B. Perron (2012). Beyond panel unit root tests: Using multiple testing to determine the non stationarity properties of individual series in a panel. *Journal of Econometrics* 169, 29–33.
- Palm, F. C., S. Smeeke, and J.-P. Urbain (2008). Bootstrap unit root tests: comparison and extensions. *Journal of Time Series Analysis* 29, 371–401.
- Palm, F. C., S. Smeeke, and J.-P. Urbain (2011). Cross-sectional dependence robust block bootstrap panel unit root tests. *Journal of Econometrics* 163, 85–104.
- Paparoditis, E. and D. N. Politis (2003). Residual-based block bootstrap for unit root testing. *Econometrica* 71, 813–855.
- Pedroni, P., T. J. Vogelsang, M. Wagner, and J. Westerlund (2008). Robust unit root and cointegration rank tests for time series panels. Working paper.
- Pesaran, M. H. (2012). On the interpretation of panel unit root tests. *Economics Letters* 116, 545–546.
- Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.
- Phillips, P. C. B. (1988). Regression theory for near-integrated time series. *Econometrica* 56, 1021–1043.
- Sensier, M. and D. Van Dijk (2004). Testing for volatility changes in U.S. macroeconomic time series. *Review of Economics and Statistics* 86, 833–839.
- Shao, X. (2010). The dependent wild bootstrap. *Journal of the American Statistical Association* 105, 218–235.
- Shao, X. (2011). A bootstrap-assisted spectral test of white noise under unknown dependence. *Journal of Econometrics* 162, 213–224.

- Smeeke, S. (2011). Bootstrap sequential tests to determine the stationary units in a panel. METEOR Research Memorandum 11/003, Maastricht University.
- Smeeke, S. (2013). Detrending bootstrap unit root tests. *Econometric Reviews* 32, 869–891.
- Smeeke, S. and A. M. R. Taylor (2012). Bootstrap union tests for unit roots in the presence of nonstationary volatility. *Econometric Theory* 28, 422–456.
- Smeeke, S. and J.-P. Urbain (2014). On the applicability of the sieve bootstrap in time series panels. *Oxford Bulletin of Economics and Statistics* 76, 139–15.
- Stock, J. H. and M. W. Watson (2003). Has the business cycle changed and why? In M. Gertler and K. Rogoff (Eds.), *NBER Macroeconomics Annual 2002, Volume 17*, Chapter 4, pp. 159–230. MIT Press.
- Taylor, J. B. (2011). Macroeconomic lessons from the Great Deviation. In D. Acemoglu and M. Woodford (Eds.), *NBER Macroeconomics Annual 2010, Volume 25*, Chapter 7, pp. 387–395. University of Chicago Press.
- Westerlund, J. (2014). Heteroskedasticity robust panel unit root tests. *Journal of Business and Economic Statistics* 32, 112–135.

## A Appendix: Proofs of Main Results

In order to prove the main results, we need the following preliminary lemmas. The proofs of these lemmas are based on standard arguments; details can be found in Appendix B.

**Lemma A.1.** *Let  $\sigma_t = \sigma(t/T)$  satisfy Assumption 1(ii). Then the following hold:*

1. *For any  $h = 1, \dots, T-1$ ,  $\sum_{t=1}^{T-h} \|\sigma_{t+h} - \sigma_t\| \leq C_1 h$ , where  $C_1 < \infty$  is a constant not depending on  $T$ .*
2. *For any  $\ell = 1, \dots, T$  and  $k = \lceil T/\ell \rceil$ ,  $\max_{1 \leq m \leq k} \sum_{s=1}^{\ell} \|\sigma_{(m-1)\ell+s} - \sigma_{(m-1)\ell+1}\| = O(\ell^2/T)$ .*

**Lemma A.2.** *Let  $v_t$  satisfy Assumption 1(i) and define  $\Gamma(h) = \mathbb{E} v_t v'_{t+k}$ . Then*

$$\sum_{h=-\infty}^{\infty} \|\Gamma(h)\| < \infty \quad \text{and, as } T \rightarrow \infty, \quad T^{-1} \sum_{h=-T+1}^{T-1} |h| \|\Gamma(h)\| = o(1).$$

**Lemma A.3.** *Let  $u_t$  satisfy Assumption 1 and let  $k = \lceil T/\ell \rceil$ , where  $\ell$  satisfies Assumption 2. Then for all  $r \in [0, 1]$ , we have that  $k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} U_m U'_m \xrightarrow{p} \Omega(r)$ , where  $\Omega(r)$  is defined in (6).*



We can now prove the main results.

**PROOF OF LEMMA 1.** The proof uses the same reasoning as the proofs of Lemma 1 of Cavaliere (2005) and Lemma 2 of Cavaliere et al. (2010). Define  $\varepsilon_t = \sum_{k=0}^{\infty} (\mathbb{E}_t v_{t+k} - \mathbb{E}_{t-1} v_{t+k})$ . It then follows from Theorem 3.1 of Hansen (1992) that

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \sigma_t \varepsilon_t = M_T(r) - C_T(r) \xrightarrow{d} \int_0^r \sigma(s) dB(s) = M(r),$$

where  $C_T(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} (\sigma_t - \sigma_{t-1}) z'_t - T^{-1/2} \sigma_{\lfloor Tr \rfloor} z'_{\lfloor Tr \rfloor}$  and  $z_t = \sum_{k=1}^{\infty} \mathbb{E}_t v_{t+k}$ . As  $T^{-1/2} \sup_{1 \leq t \leq T} \|z_t\| = o_p(1)$  (Hansen, 1992, p. 496), by Lemma A.1(a) we have that

$$\begin{aligned} C_T(r) &\leq T^{-1/2} \sup_{1 \leq t \leq T} \|z_t\| \sum_{t=1}^{\lfloor Tr \rfloor} \|\sigma_t - \sigma_{t-1}\| + T^{-1/2} \sup_{1 \leq t \leq T} \|\sigma_t z'_t\| \\ &\leq T^{-1/2} \sup_{1 \leq t \leq T} \|z_t\| C + \tilde{\sigma} T^{-1/2} \sup_{1 \leq t \leq T} \|z_t\|, \end{aligned}$$

which implies that  $M_T(r) \xrightarrow{d} M(r)$ . □

**PROOF OF THEOREM 1.** We treat the three bootstrap methods in turn.

**BWB:** Define  $M_r = \lceil \lfloor Tr \rfloor / \ell \rceil$  and  $N_r = \min(\ell, \lfloor Tr \rfloor - M_r)$  such that we can write  $M_T^*(r) = T^{-1/2} \sum_{m=1}^{M_r} \sum_{s=1}^{N_r} \nu_m^* u_{(m-1)\ell+s}$ . We first show that  $M_T^*(r) = T^{-1/2} \sum_{m=1}^{M_r} \sum_{s=1}^{\ell} \nu_m^* u_{(m-1)\ell+s} + o_p(1)$  uniformly in  $r$ . For this purpose we show that

$$\begin{aligned} \mathbb{P}^* \left( \sup_r \left\| T^{-1/2} \sum_{s=N_r+1}^{\ell} \nu_{M_r}^* u_{(M_r-1)\ell+s} \right\| > \epsilon \right) &\leq \mathbb{P}^* \left( T^{-1/2} \max_{1 \leq m \leq k} \max_{1 \leq j \leq \ell} \left\| \sum_{s=j}^{\ell} \nu_m^* u_{(m-1)\ell+s} \right\| > \epsilon \right) \\ &= \mathbb{P}^* \left( k^{-1/2} \max_{1 \leq m \leq k} |\nu_m^*| b_{T,m} > \epsilon \right) \xrightarrow{p} 0, \end{aligned}$$

where  $b_{T,m} = \max_{1 \leq j \leq \ell} \ell^{-1/2} \left\| \sum_{s=1}^j u_{(m-1)\ell+s} \right\|$ . Note that, for  $2 < p < 2 + \delta$ ,

$$\mathbb{P}^* \left( k^{-1/2} \max_{1 \leq m \leq k} |\nu_m^*| b_{T,m} > \epsilon \right) \leq \sum_{m=1}^k \mathbb{P}^* \left( |\nu_m^*| > \epsilon k^{1/2} b_{T,m}^{-1} \right) \leq \sum_{m=1}^k \epsilon^{-p} k^{-p/2} b_{T,m}^p \mathbb{E}^* |\nu_m^*|^p.$$

Write  $S_{m,j} = \sum_{s=1}^j v_{(m-1)\ell+s}$  and let  $\tilde{\sigma} = \sup_{0 \leq r \leq 1} \|\sigma(r)\|$ . Then

$$b_{T,m} \leq 2\ell^{-1/2} \tilde{\sigma} \max_{1 \leq j \leq \ell} \|S_{m,j}\| + 2\ell^{-1/2} \sum_{s=1}^{\ell} \|\sigma_{(m-1)\ell+s} - \sigma_{(m-1)\ell+1}\| \|v_{(m-1)\ell+s}\|$$

and, defining  $C_1$  and  $C_2$  as finite constants not depending on  $T$ ,

$$\mathbb{E} b_{T,m}^p \leq C_1 \ell^{-p/2} \mathbb{E} \left( \max_{1 \leq j \leq \ell} \|S_{m,j}\| \right)^p + C_2 \ell^{-p/2} \mathbb{E} \left( \sum_{s=1}^{\ell} \|\sigma_{(m-1)\ell+s} - \sigma_{(m-1)\ell+1}\| \|v_{(m-1)\ell+s}\| \right)^p.$$

By the stationarity of  $\{v_t\}$  and Corollary 3 in Hansen (1991), we have that for all  $m = 1, \dots, k$ ,

$$\mathbb{E} \left( \max_{1 \leq j \leq \ell} \|S_{m,j}\| \right)^p = \mathbb{E} \left( \max_{1 \leq j \leq \ell} \|S_{1,j}\| \right)^p \leq C_3 \ell^{p/2} (\mathbb{E} |v_s|^{2+\delta})^{p/(2+\delta)},$$

where  $C_3$  is a finite constant not depending on  $T$ . Furthermore, by Minkowski's inequality, the stationarity of  $v_t$  and Lemma A.1(b), we have for all  $m = 1, \dots, k$ ,

$$\begin{aligned} & \mathbb{E} \left( \sum_{s=1}^{\ell} \|\sigma_{(m-1)\ell+s} - \sigma_{(m-1)\ell+1}\| \|v_{(m-1)\ell+s}\| \right)^p \\ & \leq \mathbb{E} \|v_t\|^p \left( \max_{1 \leq m \leq k} \sum_{s=1}^{\ell} \|\sigma_{(m-1)\ell+s} - \sigma_{(m-1)\ell+1}\| \right)^p = o(1). \end{aligned}$$

It then directly follows that for all  $m = 1, \dots, k$ ,  $\mathbb{E} b_{T,m}^p < \infty$ , and consequently that  $\mathbb{P}^* (k^{-1/2} \max_{1 \leq m \leq k} |\nu_m^*| b_{T,m} > \epsilon) \xrightarrow{p} 0$ . Therefore, defining  $U_m = \ell^{-1/2} \sum_{s=1}^{\ell} u_{(m-1)\ell+s}$  and  $U_T^*(r) = k^{-1/2} \sum_{m=1}^{M_r} U_m \nu_m^*$ , it then follows directly that  $M_T^*(r) = U_T^*(r) + o_p(1)$  uniformly in  $r$ .

We can then apply the same reasoning as in Cavaliere et al. (2010, proof of Lemma A.5). Conditionally on  $\{U_m\}_{m=1}^k$ ,  $S_T^*(r)$  is a Gaussian process with independent increments and covariance  $\mathbb{E}^*[U_T^*(r)U_T^*(r)'] = k^{-1} \sum_{m=1}^{M_r} U_m U_m'$ . Therefore, the result follows if  $k^{-1} \sum_{m=1}^{M_r} U_m U_m' \xrightarrow{p} \Omega(r)$  uniformly for all  $r \in [0, 1]$ , with  $\Omega(r)$  as defined in (6). However, as  $k^{-1} \sum_{m=1}^{M_r} U_m U_m'$  is monotonically increasing in  $r$  and the limiting function is continuous in  $r$ , uniform convergence is implied by pointwise convergence (cf. Hansen, 2000, proof of Lemma A.10). Now note that  $k^{-1} \sum_{m=1}^{M_r} U_m U_m' = k^{-1} \sum_{m=1}^{\lfloor Tr \rfloor} U_m U_m' + O_p(\ell^{-1})$ , and, by Lemma A.3,  $k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} U_m U_m' \xrightarrow{p} \Omega(r)$  for each  $r \in [0, 1]$ . This completes the proof of (4) for the BWB.

**DWB:** By virtue of the Cholesky decomposition we have  $\xi_t^* = \sum_{s=0}^{\ell-1} w_{t,s} z_{t-s}^*$ , where  $\{z_t^*\}_{t=1}^T$  are i.i.d.  $N(0, 1)$ ,  $\sum_{s=0}^{\ell-1} w_{t,s} w_{t+h,s} = K(\frac{h}{\ell})$  for  $0 \leq h \leq \ell - 1$ , and  $w_{t,s} = 0$  for  $s \geq t$ . Then

$$M_T^*(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} u_t \sum_{s=t-\ell+1}^t w_{t,t-s} z_s^* = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} z_t^* \sum_{s=t}^{\lfloor Tr \rfloor \wedge t+\ell-1} w_{t,s-t} u_s.$$

Now, by defining  $\tilde{U}_{r,t} = \sum_{s=t}^{\lfloor Tr \rfloor \wedge t+\ell-1} w_{t,s-t} u_s$  for  $1 \leq t \leq T$ , we can conclude that, conditionally on the original sample,  $M_T^*(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} z_t^* \tilde{U}_{r,t}$  is a Gaussian process with

independent increments. As in the proof of the BWB it then remains to show that

$$\mathbb{E}^* M_T^*(r) M_T^*(r)' = T^{-1} \sum_{t=1}^{\lfloor Tr \rfloor} \tilde{U}_{r,t} \tilde{U}_{r,t}' = T^{-1} \sum_{s=1}^{\lfloor Tr \rfloor} \sum_{t=1}^{\lfloor Tr \rfloor} u_s u_t' K\left(\frac{s-t}{\ell}\right) \xrightarrow{p} \int_0^r \Omega(s) \, ds$$

uniformly in  $r$ . However, as this is a monotonically increasing process in  $r$  and the limit process is continuous, showing pointwise convergence again suffices. It follows directly from Theorem 2.1 of De Jong and Davidson (2000) that  $T^{-1} \sum_{s=1}^{\lfloor Tr \rfloor} \sum_{t=1}^{\lfloor Tr \rfloor} u_s u_t' K\left(\frac{s-t}{\ell}\right) \xrightarrow{p} \Omega_T(r)$ , as  $T^{-1/2} u_t$  satisfies their Assumption 2, and Assumptions 1 and 3 are also trivially satisfied. Note that they only prove the result for  $r = 1$ , but the proof can be straightforwardly adapted for any  $r \in [0, 1]$ . As  $\Omega_T(r) \rightarrow \Omega(r)$  this completes the proof of (4) for the DWB.

**AWB:** We may write  $\xi_t^* = \sum_{s=1}^t \gamma^{t-s} v_s^* = (1 - \gamma^2) \sum_{s=1}^t \gamma^{t-s} z_s^*$ , and therefore we have that

$$M_T^*(r) = T^{-1/2} (1 - \gamma^2) \sum_{t=1}^{\lfloor Tr \rfloor} \hat{u}_t \sum_{s=1}^t \gamma^{t-s} z_s^* = T^{-1/2} (1 - \gamma^2) \sum_{t=1}^{\lfloor Tr \rfloor} z_t^* \check{U}_{r,t},$$

where  $\check{U}_{r,t} = \sum_{s=t}^{\lfloor Tr \rfloor} \gamma^{s-t} u_s$ . As for the DWB, conditionally on the sample  $M_T^*(r)$  is a Gaussian process with independent increments. Therefore it remains to show that  $\mathbb{E}^* M_T^*(r) M_T^*(r)' = T^{-1} \sum_{s=1}^{\lfloor Tr \rfloor} \sum_{t=1}^{\lfloor Tr \rfloor} u_s u_t' \gamma^{(s-t)/\ell} \xrightarrow{p} \Omega(r)$  for all  $r$ . As the kernel  $K(x) = \theta^{|x|}$  satisfies Assumption 1 of De Jong and Davidson (2000), the proof of (4) is again completed by their Theorem 2.1 which provides the required convergence in probability to  $\Omega(r)$ .  $\square$

**PROOF OF THEOREM 2.** For part (a) we need to derive the limit distributions of  $\tau_{gm}$  and  $\tau_{gm}^*$  under local alternatives. The limit distribution of  $\tau_{gm}$  follows straightforwardly from the established invariance principle in Lemma 1. In particular (see e.g. Phillips, 1987, 1988), it follows directly from Lemma 1 that  $T^{-1/2} x_{\lfloor Tr \rfloor} \xrightarrow{p} M_c(r)$ , and using the continuous mapping theorem one finds that  $T^{-1/2} y_{\lfloor Tr \rfloor}^d \xrightarrow{d} M_c^d(r)$  and consequently that

$$\tau_{gm} \xrightarrow{d} \frac{1}{N} \sum_{i=1}^N \frac{M_{c,i}^d(1)^2 - M_{c,i}^d(0)^2 - \bar{\sigma}_{i,i}}{2 \int_0^1 M_{c,i}^d(r)^2 \, dr}. \quad (\text{A.1})$$

For the bootstrap methods we first need to adapt the bootstrap invariance principle of Theorem 1 to acknowledge the fact that now  $\{\hat{u}_t\}$  is used to construct the bootstrap errors rather than  $\{u_t\}$ . That is, we need the bootstrap invariance principle

$$X_T^*(r) = T^{-1/2} x_{\lfloor Tr \rfloor}^* = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \xi_t^* \hat{u}_t \xrightarrow{d^*} M(r). \quad (\text{A.2})$$

For  $\tau_{gm}^*$  it then follows straightforwardly that

$$\tau_{gm}^* \xrightarrow{d^*} \frac{1}{N} \sum_{i=1}^N \frac{M_i^d(1)^2 - M_i^d(0)^2 - \bar{\sigma}_{i,i}}{2 \int_0^1 M_i^d(r)^2 dr}, \quad (\text{A.3})$$

by the bootstrap invariance principle (A.2) and the continuous mapping theorem. We now again treat the three bootstrap methods in turn to prove (A.2).

**BWB:** We follow the same approach as in the proof of Theorem 1 to show that  $X_T^*(r) = T^{-1/2} \sum_{m=1}^{M_r} \sum_{s=1}^{\ell} \nu_m^* \hat{u}_{(m-1)\ell+s} + o_p(1)$  uniformly in  $r$ . Define  $\hat{b}_{T,m} = \max_{1 \leq j \leq \ell} \left\| \ell^{-1/2} \sum_{s=j}^{\ell} \hat{u}_{(m-1)\ell+s} \right\|$  and note that as  $\hat{u}_t = u_t - (\hat{\rho} - \rho)x_{t-1} + (\hat{\rho} - 1)\bar{x}$ , we have that

$$\hat{b}_{T,m} \leq b_{T,m} + \max_{1 \leq j \leq \ell} \|\hat{\rho} - \rho\| \left\| \ell^{-1/2} \sum_{s=j}^{\ell} x_{(m-1)\ell+s-1} \right\| + \|(\hat{\rho} - 1)\bar{x}\|.$$

For the third term  $\|(\hat{\rho} - 1)\bar{x}\| = O_p(T^{-1/2})$ , while for the second term we have that

$$\begin{aligned} \max_{1 \leq j \leq \ell} \left\| \ell^{-1/2} \sum_{s=j}^{\ell} x_{(m-1)\ell+s-1} \right\| &\leq 2 \max_{1 \leq j \leq \ell} \left\| \ell^{-1/2} \sum_{s=1}^j x_{(m-1)\ell+s-1} \right\| \\ &\leq 2 \max_{1 \leq j \leq \ell} \left\| \ell^{-1/2} j x_{(m-1)\ell} + \ell^{-1/2} \sum_{s=2}^j s u_{(m-1)\ell+s-1} \right\| \\ &\leq 2\ell^{1/2} \|x_{(m-1)\ell}\| + 2\tilde{\sigma}\ell^{1/2} \sum_{s=1}^{\ell} \|v_{(m-1)\ell+s}\| = O_p(\ell^{1/2}T^{1/2}) + O_p(\ell^{3/2}). \end{aligned}$$

As  $\|\hat{\rho} - \rho\| = O_p(T^{-1})$ , it then follows that  $k^{-p/2} \sum_{m=1}^k \hat{b}_{T,m}^p = k^{-p/2} \sum_{m=1}^k b_{T,m}^p + o_p(1)$  and consequently, along the lines of the proof of Theorem 1, that

$$\mathbb{P}^* \left( \sup_r \left\| T^{-1/2} \sum_{s=N_r+1}^{\ell} \nu_m^* \hat{u}_{(m-1)\ell+s} \right\| > \epsilon \right) = \mathbb{P}^* \left( k^{-1/2} \max_{1 \leq m \leq k} |\nu_m^*| \hat{b}_{T,m} > \epsilon \right) \xrightarrow{p} 0.$$

Define  $\hat{U}_m = \ell^{-1/2} \sum_{s=1}^{\ell} \hat{u}_{(m-1)\ell+s}$  and  $S_T^*(r) = k^{-1/2} \sum_{m=1}^{M_r} \hat{U}_m \nu_m^*$ , such that conditionally on  $\{\hat{U}_m\}_{m=1}^k$ ,  $S_T^*(r)$  is a Gaussian process with independent increments and covariance  $\mathbb{E}^*[S_T^*(r)S_T^{*'}(r)] = k^{-1} \sum_{m=1}^{M_r} \hat{U}_m \hat{U}_m'$ . The result then follows if we can show that  $k^{-1} \sum_{m=1}^{M_r} \hat{U}_m \hat{U}_m' \xrightarrow{p} k^{-1} \sum_{m=1}^{M_r} U_m U_m'$  for all  $r \in [0, 1]$ , as uniform convergence is again implied by pointwise convergence.

Note that

$$\begin{aligned}
& k^{-1} \left\| \sum_{m=1}^{M_r} \hat{U}_m \hat{U}'_m - \sum_{m=1}^{M_r} U_m U'_m \right\| \leq 2(k\ell)^{-1} \|\hat{\rho} - \rho\| \sum_{m=1}^{M_r} \left\| \sum_{s=1}^{\ell} u_{(m-1)\ell+s} \right\| \left\| \sum_{s=1}^{\ell} x_{(m-1)\ell+s-1} \right\| \\
& + 2(k\ell)^{-1} \|\hat{\rho} - 1\| \|\bar{x}\| \sum_{m=1}^{M_r} \left\| \sum_{s=1}^{\ell} u_{(m-1)\ell+s} \right\| + (k\ell)^{-1} \|\hat{\rho} - \rho\|^2 \sum_{m=1}^{M_r} \left\| \sum_{s=1}^{\ell} x_{(m-1)\ell+s} \right\|^2 \\
& + 2(k\ell)^{-1} \|\hat{\rho} - 1\| \|\hat{\rho} - \rho\| \|\bar{x}\| \sum_{m=1}^{M_r} \left\| \sum_{s=1}^{\ell} x_{(m-1)\ell+s} \right\| + M_r (k\ell^{-1} \|\hat{\rho} - 1\|^2 \|\bar{x}\|^2).
\end{aligned}$$

As  $\left\| \sum_{s=1}^{\ell} x_{(m-1)\ell+s} \right\| = O_p(\ell T^{-1/2})$ ,  $\left\| \sum_{s=1}^{\ell} u_{(m-1)\ell+s} \right\| = O_p(\ell^{1/2})$ ,  $\|\hat{\rho} - \rho\| = O_p(T^{-1})$  and  $\|\bar{x}\| = O_p(T^{1/2})$ , it follows straightforwardly that  $k^{-1} \sum_{m=1}^{M_r} \hat{U}_m \hat{U}'_m = k^{-1} \sum_{m=1}^{M_r} U_m U'_m + O_p(k^{-1/2})$ , and (A.2) now follows directly from the proof of Theorem 1.

**DWB:** As for Theorem 1 we can write  $X_T^*(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} z_t^* \sum_{s=t}^{\lfloor Tr \rfloor \wedge t+\ell-1} w_{t,s-t} \hat{u}_s$ . Then by the same reasoning as for the BWB, for (A.2) to hold it remains to show that

$$\mathbb{E}^* X_T^*(r) X_T^*(r)' = T^{-1} \sum_{s=1}^{\lfloor Tr \rfloor} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{u}_s \hat{u}'_t K\left(\frac{s-t}{\ell}\right) = T^{-1} \sum_{s=1}^{\lfloor Tr \rfloor} \sum_{t=1}^{\lfloor Tr \rfloor} u_s u'_t K\left(\frac{s-t}{\ell}\right) + o_p(1)$$

for all  $r \in [0, 1]$ . Note that  $\sum_{s=1}^{\lfloor Tr \rfloor} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{u}_s \hat{u}'_t K\left(\frac{s-t}{\ell}\right) = \sum_{h=0}^{\ell} K\left(\frac{h}{\ell}\right) \sum_{t=1}^{\lfloor Tr \rfloor-h} \hat{u}_t \hat{u}'_{t+h} + \sum_{h=1}^{\ell} K\left(\frac{h}{\ell}\right) \sum_{t=1}^{\lfloor Tr \rfloor-h} \hat{u}_{t+h} \hat{u}'_t$ . By applying the Cauchy-Schwartz inequality we have that

$$\begin{aligned}
& T^{-1} \left\| \sum_{h=0}^{\ell} K\left(\frac{h}{\ell}\right) \sum_{t=1}^{\lfloor Tr \rfloor-h} (\hat{u}_t \hat{u}'_{t+h} - u_t u'_{t+h}) \right\| \\
& \leq 2T^{-1} \|\hat{\rho} - \rho\| \sum_{h=0}^{\ell} K\left(\frac{h}{\ell}\right) \left( \sum_{t=1}^{\lfloor Tr \rfloor-h} \|u_t\|^2 \right)^{1/2} \left( \sum_{t=1}^{\lfloor Tr \rfloor-h} \|x_t\|^2 \right)^{1/2} \\
& + 2T^{-1} \|\hat{\rho} - 1\| \|\bar{x}\| \sum_{h=0}^{\ell} K\left(\frac{h}{\ell}\right) \sum_{t=1}^{\lfloor Tr \rfloor-h} \|u_t\| + T^{-1} \|\hat{\rho} - \rho\|^2 \sum_{h=0}^{\ell} K\left(\frac{h}{\ell}\right) \sum_{t=1}^{\lfloor Tr \rfloor-h} \|x_t\|^2 \\
& + 2T^{-1} \|\hat{\rho} - 1\| \|\hat{\rho} - \rho\| \|\bar{x}\| \sum_{h=0}^{\ell} K\left(\frac{h}{\ell}\right) \sum_{t=1}^{\lfloor Tr \rfloor-h} \|x_t\| + \|\hat{\rho} - 1\|^2 \|\bar{x}\|^2 \sum_{h=0}^{\ell} K\left(\frac{h}{\ell}\right).
\end{aligned}$$

It then follows that  $T^{-1} \sum_{s=1}^{\lfloor Tr \rfloor} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{u}_s \hat{u}'_t K\left(\frac{s-t}{\ell}\right) = T^{-1} \sum_{s=1}^{\lfloor Tr \rfloor} \sum_{t=1}^{\lfloor Tr \rfloor} u_s u'_t K\left(\frac{s-t}{\ell}\right) + O_p(\ell T^{-1/2})$ .

**AWB:** Analogously to the proof for the DWB we need to show that

$$\mathbb{E}^* Y_T^*(r) Y_T^*(r)' = T^{-1} \sum_{s=1}^{\lfloor Tr \rfloor} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{u}_s \hat{u}_t' \gamma^{(s-t)/\ell} = T^{-1} \sum_{s=1}^{\lfloor Tr \rfloor} \sum_{t=1}^{\lfloor Tr \rfloor} u_s u_t' \gamma^{(s-t)/\ell} + o_p(1)$$

for all  $r \in [0, 1]$ . Using that  $\gamma = \theta^{1/\ell}$ , we can write

$$\begin{aligned} & T^{-1} \left\| \sum_{h=0}^{\lfloor Tr \rfloor - 1} \theta^{\frac{h}{\ell}} \sum_{t=1}^{\lfloor Tr \rfloor - h} (\hat{u}_t \hat{u}_{t+h}' - u_t u_{t+h}') \right\| \\ & \leq 2T^{-1} \|\hat{\rho} - \rho\| \sum_{h=0}^{\lfloor Tr \rfloor - 1} \theta^{\frac{h}{\ell}} \left( \sum_{t=1}^{\lfloor Tr \rfloor - h} \|u_t\|^2 \right)^{1/2} \left( \sum_{t=1}^{\lfloor Tr \rfloor - h} \|x_t\|^2 \right)^{1/2} \\ & \quad + 2T^{-1} \|\hat{\rho} - 1\| \|\bar{x}\| \sum_{h=0}^{\lfloor Tr \rfloor - 1} \theta^{\frac{h}{\ell}} \sum_{t=1}^{\lfloor Tr \rfloor - h} \|u_t\| + T^{-1} \|\hat{\rho} - \rho\|^2 \sum_{h=0}^{\ell} \theta^{\frac{h}{\ell}} \sum_{t=1}^{\lfloor Tr \rfloor - h} \|x_t\|^2 \\ & \quad + 2T^{-1} \|\hat{\rho} - 1\| \|\hat{\rho} - \rho\| \|\bar{x}\| \sum_{h=0}^{\lfloor Tr \rfloor - 1} \theta^{\frac{h}{\ell}} \sum_{t=1}^{\lfloor Tr \rfloor - h} \|x_t\| + \|\hat{\rho} - 1\|^2 \|\bar{x}\|^2 \sum_{h=0}^{\lfloor Tr \rfloor - 1} \theta^{\frac{h}{\ell}}. \end{aligned}$$

As  $\sum_{h=0}^{\lfloor Tr \rfloor - 1} \theta^{\frac{h}{\ell}} = \frac{1 - \theta^{\lfloor Tr \rfloor / \ell}}{1 - \theta^{1/\ell}} = -\frac{\ell}{\ln \theta} + o(\ell)$ , (A.2) follows straightforwardly.

For part (b), first note that it follows analogously to Lemma 6 of Palm et al. (2011) that  $\tau_{gm}$  diverges to  $-\infty$  under  $H_1$ . Before we can derive the bootstrap distribution, we need a reparametrization of our model that is better suited to deal with fixed alternatives. Let us assume without loss of generality that for the first  $N_1$  units  $\rho_i < 1$ , while for the remaining units  $\rho_i = 1$ . Then define  $\tilde{\rho} = \text{diag}(\tilde{\rho}_1, \dots, \tilde{\rho}_N)$ , where  $\tilde{\rho}_i = \text{plim } \hat{\rho}_i$  for  $i = 1, \dots, N_1$ , where  $\hat{\rho}_i$  is the estimator used in step 1 of Algorithm 1, and  $\tilde{\rho}_i = 1$  for  $i = N_1 + 1, \dots, N$ . Next let  $\tilde{u}_t = u_t - (\tilde{\rho} - \rho)x_{t-1}$  such that  $x_t = \tilde{\rho}x_{t-1} + \tilde{u}_t$ . Note that we can write  $\tilde{u}_t = \tilde{\sigma}_t \tilde{v}_t$  such that it satisfies Assumption 1. It then straightforwardly follows that  $T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \tilde{u}_t \xrightarrow{d} \tilde{M}(r)$ , where  $\tilde{M}(r)$  is defined implicitly.

To derive the bootstrap limit distribution we can then simply reproduce the proof of (A.2). The only modification required is when the transition is made in the variance estimators involving  $\hat{u}_t$  to the unobservable  $u_t$ . In that proof simply replace  $u_t$  by  $\tilde{u}_t$  and  $\rho$  by  $\tilde{\rho}$  and note that  $\left\| \sum_{s=1}^{\ell} x_{(m-1)\ell+s} \right\| = O_p(\ell^{1/2})$ ,  $\left\| \sum_{s=1}^{\ell} \tilde{u}_{(m-1)\ell+s} \right\| = O_p(\ell^{1/2})$ ,  $\|\hat{\rho} - \tilde{\rho}\| = O_p(T^{-1/2})$  and  $\|\bar{x}\| = O_p(T^{-1/2})$ , then the proof goes through straightforwardly. We then recover (A.3) with  $\tilde{M}(r)$  replacing  $M(r)$ , and the asymptotic power result follows directly.<sup>8</sup>  $\square$

<sup>8</sup>The bootstrap may appear to reproduce a different distribution under the alternative than under the null, but this is not so. It is purely an artefact of our specific parametrization; if a DF-type estimator of  $\hat{\rho}_i$  is used in step 1 of the algorithm,  $\rho$  cannot be estimated consistently. This is no problem as  $\tilde{\rho}$  can be estimated consistently, and the properties of  $\tilde{u}_t$  under the alternative are identical to  $u_t$  under the null. The exact value of  $\rho$  under a fixed alternative therefore does not matter (also see Remark 7 of Palm et al., 2011).

## B Appendix: Proofs of Preliminary Lemmas

PROOF OF LEMMA A.1. For the proof of (a), suppose that the  $M < \infty$  discontinuities are located at  $0 < x_1, \dots, x_M < 1$ , and use these points to partition  $[0, 1]$  where we add the points  $x_0 = 0$  and  $x_{M+1} = 1$ . Let  $T_i = \lfloor x_i T \rfloor$ , then

$$\begin{aligned} \sum_{t=1}^{T-h} \|\sigma_{t+h} - \sigma_t\| &= \sum_{i=1}^M \sum_{t=T_{i-1}+1}^{T_i \wedge T-h} \|\sigma_{t+h} - \sigma_t\| \\ &= \sum_{i=1}^{M+1} \sum_{t=T_{i-1}+1}^{T_i-h} \|\sigma_{t+h} - \sigma_t\| + \sum_{i=1}^{M+1} \sum_{t=T_i-h+1}^{T_i \wedge T-h} \|\sigma_{t+h} - \sigma_t\| = R_{1,h} + R_{2,h}. \end{aligned}$$

Note that by construction in the terms contained in  $R_{1,h}$  no jumps occur. Therefore these terms can be bounded using the Lipschitz condition. In particular, for some  $C < \infty$ , we have that

$$R_{1,h} = \sum_{i=1}^M \sum_{t=T_{i-1}+1}^{T_i-h} \left\| \sigma\left(\frac{t+h}{T}\right) - \sigma\left(\frac{t}{T}\right) \right\| \leq \sum_{i=1}^M (T_i - T_{i-1} + 1 - h) C \frac{h}{T} = C(T-h+1) \frac{h}{T} \leq Ch.$$

On the other hand, all discontinuities occur in  $R_{2,T}$ . Yet, as they are finite,  $\tilde{\sigma} = \sup_{0 \leq r \leq 1} \|\sigma(r)\| < \infty$ , and for each  $i$  the inner sum contains at most  $h$  items, we have that

$$R_{2,T} \leq \sum_{i=1}^M \sum_{t=T_i-h+1}^{T_i \wedge T-h} 2\tilde{\sigma} \leq 2\tilde{\sigma} Mh.$$

The result then follows with  $C_1 = C + 2\tilde{\sigma}M$ .

For (b) note that given the finite number of discontinuities and the right-continuity of  $\sigma(\cdot)$ , there is a  $T^*$  such that for all  $T > T^*$  no break occurs between  $\sigma(r + \eta_T)$  and  $\sigma(r)$  for any sequence  $\{\eta_T\}_{T=1}^\infty$  such that  $\eta_T \rightarrow 0$  as  $T \rightarrow \infty$ . Then taking  $\eta_T = \ell/T$ , there is a  $T^*$  such that for all  $T > T^*$ :

$$\begin{aligned} \max_{1 \leq m \leq k} \sum_{s=1}^{\ell} \|\sigma_{(m-1)\ell+s} - \sigma_{(m-1)\ell+1}\| &\leq \sup_{0 \leq r \leq 1-\ell/T} \sum_{s=1}^{\ell} \left\| \sigma\left(r + \frac{s}{T}\right) - \sigma\left(r + \frac{1}{T}\right) \right\| \\ &\leq \sup_{0 \leq r \leq 1-\ell/T} \sum_{s=1}^{\ell} C \frac{s-1}{T} \leq C\ell^2/T. \quad \square \end{aligned}$$

PROOF OF LEMMA A.2. By Proposition 3.1(b) of Lahiri (2003) and Assumption 1 we have that, as  $1/p - 1/(2 + \delta) < \delta/(2 + \delta)$ ,

$$\sum_{h=-\infty}^{\infty} \|\Gamma(h)\| \leq 8C_1(N) (\mathbb{E} \|v_t\|^p) \sum_{h=-\infty}^{\infty} \alpha(h)^{\delta/(2+\delta)} \leq C_2(N) \sum_{h=-\infty}^{\infty} \alpha(h)^{1/p-1/(2+\delta)} < \infty,$$

where  $C_1(N)$  and  $C_2(N)$  are constants not depending on  $T$ . Define  $Q_T$  such that  $Q_T \rightarrow \infty$  and  $Q_T = o(T)$  as  $T \rightarrow \infty$ . Then

$$\begin{aligned} T^{-1} \sum_{h=0}^{T-1} h \|\Gamma(h)\| &= T^{-1} \sum_{h=0}^{Q_T} m \|\Gamma(h)\| + T^{-1} \sum_{h=Q_T+1}^{T-1} h \|\Gamma(h)\| \\ &\leq Q_T T^{-1} \sum_{h=0}^{Q_T} \|\Gamma(h)\| + \sum_{h=Q_T+1}^{T-1} \|\Gamma(h)\| = o(1), \end{aligned}$$

as  $\sum_{h=0}^{Q_T} \|\Gamma(h)\| = O(1)$  by the first part of the lemma and  $\sum_{h=Q_T+1}^{T-1} \|\Gamma(h)\| = o(1)$  by Theorem 2.25 of Davidson (2002).  $\square$

PROOF OF LEMMA A.3. Note that

$$k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} U_m U'_m - \Omega(r) = k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} (U_m U'_m - \mathbb{E} U_m U'_m) + \left( k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} \mathbb{E} U_m U'_m - \Omega(r) \right).$$

We first show that  $k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} (U_m U'_m - \mathbb{E} U_m U'_m)$  using an adaptation of Theorem 3.1 of Lahiri (2003) to the case of nonstationary volatility, and for the sum running up to  $\lfloor kr \rfloor$  instead of  $k$ . Note that Define  $V_{m,T} = U_m U'_m I(\|U_m\| < (T/\ell)^{1/8})$ . Also let  $W_{m,T} = U_m U'_m - V_{m,T}$ , then

$$k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} (U_m U'_m - \mathbb{E} U_m U'_m) = k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} (V_{m,T} - \mathbb{E} V_{m,T}) + k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} (W_{m,T} - \mathbb{E} W_{m,T}).$$

Define  $\|f\|_\infty = \sup \{|f(x)| : x \in \mathbb{R}^N\}$ . By applying Lemma 3.1 of Lahiri (2003) element by element, with  $f_{ij,T}(x) = x_i x_j I(\|x\| < (T/\ell)^{1/8})$ , we have that

$$\begin{aligned} \mathbb{E} \left\| k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} (V_{m,T} - \mathbb{E} V_{m,T}) \right\|^2 &\leq k^{-2} C \|f\|_\infty^2 k \left[ 1 + \sum_{1 \leq m \leq \lfloor kr \rfloor} \alpha(m\ell) \right] \\ &= C(T/\ell)^{1/2} k^{-1} \left[ 1 + \sum_{1 \leq m \leq \lfloor kr \rfloor} \alpha(m\ell) \right] \\ &\leq C k^{-1/2} \left[ 1 + \sum_{m=1}^{\lfloor kr \rfloor} \alpha(m) \right] = o(1), \end{aligned}$$

as  $C < \infty$  does not depend on  $T$ . Hence, by the Markov inequality we have that  $k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} (V_{m,T} -$



$\mathbb{E} V_{m,T}) = o_p(1)$ . Furthermore,

$$\mathbb{E} \left\| k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} (W_{m,T} - \mathbb{E} W_{m,T}) \right\| \leq k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} \mathbb{E} \|W_{m,T} - \mathbb{E} W_{m,T}\| \leq 2k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} \mathbb{E} \|W_{m,T}\|.$$

As  $\|W_{m,T}\| \leq \|U_m\|^2$  for all  $T$ , then as  $U_m = O_p(1)$ , we can conclude by the extended dominated convergence theorem that  $k^{-1} \sum_{m=1}^k \mathbb{E} \|W_{m,T}\| = o(1)$  and hence that  $k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} (W_{m,T} - \mathbb{E} W_{m,T}) = o_p(1)$  by the Markov inequality. This completes the first part of the proof.

Next we show that  $k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} \mathbb{E} U_m U'_m \xrightarrow{p} \Omega(r)$  for  $r \in [0, 1]$ . First note that

$$\left\| k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} \mathbb{E} U_m U'_m - \Omega(r) \right\| \leq \left\| k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} \mathbb{E} U_m U'_m - \Omega_T(r) \right\| + \|\Omega_T(r) - \Omega(r)\|$$

Then, noting that  $\tilde{\sigma} = \max_{1 \leq t \leq T} \|\sigma_t\| < \infty$ , we have

$$\begin{aligned} & \left\| k^{-1} \sum_{m=1}^{\lfloor kr \rfloor} \mathbb{E} U_m U'_m - \Omega_T \right\| \\ &= \left\| T^{-1} \sum_{m=1}^{\lfloor kr \rfloor} \sum_{s_1=1}^{\ell} \sum_{s_2=1}^{\ell} \sigma_{(m-1)\ell+s_1} \Gamma(s_1 - s_2) \sigma'_{(m-1)\ell+s_2} - T^{-1} \sum_{t_1=1}^{\lfloor Tr \rfloor} \sum_{t_2=1}^{\lfloor Tr \rfloor} \sigma_{t_1} \Gamma(t_1 - t_2) \sigma'_{t_2} \right\| \\ &\leq T^{-1} \sum_{m_1=1}^k \sum_{m_2=1}^k (1 - \delta_{m_1 m_2}) \sum_{s_1=1}^{\ell} \sum_{s_2=1}^{\ell} \left\| \sigma_{(m_1-1)\ell+s_1} \Gamma([m_1 - m_2]\ell + [s_1 - s_2]) \sigma'_{(m_2-1)\ell+s_2} \right\| \\ &\leq T^{-1} \tilde{\sigma}^2 \sum_{m_1=1}^k \sum_{m_2=1}^k (1 - \delta_{m_1 m_2}) \sum_{s_1=1}^{\ell} \sum_{s_2=1}^{\ell} \|\Gamma([m_1 - m_2]\ell + [s_1 - s_2])\| \\ &\leq 2\tilde{\sigma}^2 \frac{k}{T} \sum_{m=1}^{\ell} m \|\Gamma(m)\| + 2\tilde{\omega}^2 \sum_{m=\ell+1}^{T-1} \frac{T-m}{T} \|\Gamma(m)\| = o(1), \end{aligned}$$

where the final equality follows from Lemma A.2.

To finalize the proof, define  $\Sigma_{T,0}(r) = T^{-1} \sum_{t=1}^{\lfloor Tr \rfloor} \sigma_t \sigma'_t$  and  $\Upsilon_T(r) = T^{-1} \sum_{h=1}^{\lfloor Tr \rfloor - 1} \sum_{t=h+1}^{\lfloor Tr \rfloor} \sigma_t \Gamma(h) \sigma'_{t-h}$ , such that  $\Omega_T(r) = \Sigma_{T,0}(r) + \Upsilon_T(r) + \Upsilon_T(r)'$ . Now let  $\tilde{\Upsilon}_T(r) = T^{-1} \sum_{h=1}^{\lfloor Tr \rfloor - 1} \sum_{t=1}^{\lfloor Tr \rfloor} \sigma_t \Gamma(h) \sigma'_t$

and note that, using Lemmas A.1 and A.2,

$$\begin{aligned}
\left\| \Upsilon_T(r) - \tilde{\Upsilon}_T(r) \right\| &\leq T^{-1} \sum_{h=1}^{T-1} \|\Gamma(h)\| \left[ \sum_{t=h+1}^T \|\sigma_t\| \|\sigma_t - \sigma_{t-h}\| + \sum_{t=1}^h \|\sigma_t\|^2 \right] \\
&\leq \tilde{\sigma} T^{-1} \sum_{h=1}^{T-1} \|\Gamma(h)\| \left[ \sum_{t=h+1}^T \|\sigma_t - \sigma_{t-h}\| + h\tilde{\sigma} \right] \\
&\leq \tilde{\sigma}(C_1 + \tilde{\sigma}) T^{-1} \sum_{h=1}^{T-1} h \|\Gamma(h)\| = o(1),
\end{aligned}$$

which confirms that  $\Omega_T(r) \rightarrow \Omega(r)$ . □